

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出 願 年 月 日 2 0 0 3 年 8 月 6 日
Date of Application:

出 願 番 号 特 願 2 0 0 3 - 2 0 6 1 6 8
Application Number:
ST. 10/C] : [J P 2 0 0 3 - 2 0 6 1 6 8]

願 人
Applicant(s): 株式会社日立製作所

CERTIFIED COPY OF
PRIORITY DOCUMENT

USSN 10/712,031
MATTINGLY, STANGER, MALUR + BRUNDIGE
(703) 684-1120
TSM-34

2 0 0 3 年 1 0 月 1 5 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康 夫



BEST AVAILABLE COPY

出証番号 出証特 2 0 0 3 - 3 0 8 4 5

【書類名】 特許願

【整理番号】 K03009841A

【あて先】 特許庁長官殿

【国際特許分類】 G06F 12/00

【発明者】

 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所システム開発研究所内

 【氏名】 岩見 直子

【発明者】

 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所システム開発研究所内

 【氏名】 山本 康友

【発明者】

 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所システム開発研究所内

 【氏名】 藤本 和久

【特許出願人】

 【識別番号】 000005108

 【氏名又は名称】 株式会社 日立製作所

【代理人】

 【識別番号】 100075096

 【弁理士】

 【氏名又は名称】 作田 康夫

【手数料の表示】

 【予納台帳番号】 013088

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

【物件名】 要約書 1
【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 ストレージ装置

【特許請求の範囲】

【請求項 1】

計算機と接続されるインターフェース部と、
ファイル操作の処理を行う第一の制御部と
記憶装置に対するデータの読み書きの処理を行う第二の制御部と、
前記インターフェース部、前記第一の制御部及び前記第二の制御部とを相互に
接続する接続網とを有し、

前記インターフェース部は、前記計算機から送信されたフレームの転送先を前
記第一の制御部及び前記第二の制御部のうちいずれか一つから選択し、前記接続
網を介して前記フレームを選択した制御部へ転送することを特徴とするストレ
ージ装置。

【請求項 2】

前記第一の制御部は、前記インターフェース部から前記フレームを受信して前
記フレームで指定されたファイル操作を行う場合、前記第二の制御部を介して前
記記憶装置に対するデータの読み書きを実行することを特徴とする請求項 1 記載
のストレージ装置。

【請求項 3】

他のストレージ装置と接続する第二のインターフェース部を有し、
前記インターフェース部は、前記フレームの転送先を前記第一の制御部、前記
第二の制御部又は前記第二のインターフェース部から選択することを特徴とする
請求項 2 記載のストレージ装置。

【請求項 4】

前記第一の制御部は、前記インターフェース部から前記フレームを受信して前
記フレームで指定されたファイル操作を行う場合、前記第二のインターフェース
部を介して前記他のストレージ装置に対するデータの読み書きを実行することを
特徴とする請求項 3 記載のストレージ装置。

【請求項 5】

前記第一の制御部を複数有し、

前記インターフェース部は、前記フレームがファイル操作を要求するコマンドを含むフレームであった場合、前記複数の第一の制御部から所定の第一の制御部を選択し、選択された前記第一の制御部へ前記フレームを前記接続網を介して送信することを特徴とする請求項 1 記載のストレージ装置。

【請求項 6】

前記インターフェース部は、前記複数の第一の制御部と前記計算機から受信するフレームに含まれる識別子との対応関係に関する情報を保持し、前記フレームを受信した際に前記情報に基づいて前記フレームを転送する先の第一の制御部を決定することを特徴とする請求項 5 記載のストレージ装置。

【請求項 7】

前記フレームを受信した前記第一の制御部の指示により、前記インターフェース部及び前記第二の制御部は、前記接続網を介して前記第一の制御部の処理に関するデータを送受信することを特徴とする請求項 1 記載のストレージ装置。

【請求項 8】

前記フレームを受信した前記第一の制御部の指示により、前記インターフェース部及び前記第二のインターフェース部は、前記接続網を介して前記第一の制御部の処理に関するデータを送受信することを特徴とする請求項 4 記載のストレージ装置。

【請求項 9】

前記情報には、前記フレームを受信する前記インターフェース部が有する一つのポートに複数の前記第一の制御部が対応することを示す情報が含まれていることを特徴とする請求項 6 記載のストレージ装置。

【請求項 10】

前記インターフェース部を複数有し、前記複数のインターフェース部で受信されたフレームが前記第一の制御部に転送されることを特徴とする請求項 1 記載のストレージ装置。

【請求項 11】

更に管理部を有し、前記インターフェース部は、前記管理部の指示に従って、

前記情報の内容を変更し、前記フレームの転送先を変更することを特徴とする請求項 9 記載のストレージ装置。

【請求項 1 2】

前記インターフェース部を複数有し、

前記管理部の指示に従って、前記インターフェース部で行っていた処理を他の前記インターフェース部に引継ぐ手段を有することを特徴とする請求項 1 1 記載のストレージ装置。

【請求項 1 3】

前記管理部の指示は、前記インターフェース部の障害を該ストレージ装置が有する各装置が検出した際に行われ、前記管理部は、前記インターフェース部の障害の際に処理を引継ぐ前記他のインターフェース部の情報を有することを特徴とする請求項 1 2 記載のストレージ装置。

【請求項 1 4】

前記第一の制御部を複数有し、

前記管理部の指示に従って、前記第一の制御部で行っていた処理を他の前記第一の制御部に引継ぐ手段を有することを特徴とする請求項 1 2 記載のストレージ装置。

【請求項 1 5】

前記管理部の指示は、前記第一の制御部の障害を該ストレージ装置が有する各装置が検出した際に行われ、前記管理部は、前記第一の制御部の障害の際に処理を引継ぐ前記他の第一の制御部の情報を有することを特徴とする請求項 1 4 記載のストレージ装置。

【請求項 1 6】

前記第二の制御部は、前記第二のインターフェース部を介して前記他のストレージ装置を制御することを特徴とする請求項 4 記載のストレージ装置。

【請求項 1 7】

前記第二の制御部は、キャッシュメモリ及びディスク装置を有することを特徴とする請求項 1 から 1 6 に記載のうちいずれか一つのストレージ装置。

【請求項 1 8】

計算機と接続されるインターフェース部と、

ファイル操作の処理を行う制御部と

記憶装置に対するデータの読み書きの処理を行う第二のインターフェース部と

、
前記インターフェース部、前記第二のインターフェース部及び前記制御部とを相互に接続する接続網とを有し、

前記インターフェース部は、前記計算機から送信されたフレームの転送先を前記第一の制御部及び前記第二のインターフェース部のうちいずれか一つから選択し、前記接続網を介して前記フレームを選択した前記第一の制御部又は前記第二のインターフェース部へ転送することを特徴とするストレージ装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、計算機システムにおいて計算機によって使用されるデータを格納するストレージ装置、特にネットワーク経由でファイルアクセスされるストレージ装置に関する。

【0002】

【従来の技術】

近年、計算機で取り扱われるデータ量は飛躍的に増大し、これに応じてデータを格納するストレージ装置の大容量化が進んでいる。大規模かつ低コストな大型ストレージを実現する方法として、特許文献1に開示されているクラスタ構成のストレージシステムがある。

【0003】

クラスタ構成ストレージシステムは、比較的小構成のストレージ（以下「ストレージノード」）複数台をスイッチ等で相互に結合した構成を有する。このような構成では、ストレージノードを必要とされる記憶容量に合わせて複数結合してクラスタ構成ストレージシステムを構成することができ、小容量から大容量まで様々な容量を持つストレージシステムをスケラブルに実現できる。

【0004】

一方、ストレージ装置の種類の一つとして、ネットワークアタッチドストレージ（以下「NAS」）がある。これは、ネットワーク経由でのファイルアクセスを受け付けることが可能なストレージ装置で、実際には、ファイルシステムを提供する計算機（以下「NASサーバ」）及びディスク装置等の記憶装置から構成される。尚、「ファイル」とは、ファイルシステム等でユーザが認識するデータの一塊を指す。

NASは、使用者がファイルアクセスを直接利用できるため使いやすいストレージ装置である。しかし、NASへアクセスするためのプロトコル（N F S等）が、TCP/IP(Transmission Control Protocol/Internet Protocol)の使用を前提としているため、NASへのアクセス頻度が上がるにつれ、NASにおけるTCP/IPの処理負荷が問題となってきた。この問題を解決する技術として、TCP/IPの処理をNASサーバから分離して単独のハードウェア（以下「TCP/IP処理部」）とし、TCP/IP処理部とNASサーバをスイッチで接続する技術が特許文献 2 に開示されている。特許文献 2 では、計算機が発行したコマンドは、TCP/IP処理部でTCP/IPレイヤまで処理され、その後NASサーバ部で処理される。NASサーバが記憶装置をアクセスする場合は、Fiber Channel Switching Module経由でSANに接続された記憶装置を利用する。

【 0 0 0 5 】

【特許文献 1】

米国特許 6, 2 5 6, 7 4 0 号

【特許文献 2】

P C T 国際公開 0 2 / 0 6 9 1 6 6 号公報

【 0 0 0 6 】

【発明が解決しようとする課題】

特許文献 2 に開示されている技術では、アクセス対象となるNASサーバとNASサーバに付与されたネットワークアドレスが設定されるTCP/IP処理部のポートとの関係は 1 : 1 で一意に決定される。このため、NASサーバに対するアクセス頻度が上がったり、送受するデータ量が増加すると、TCP/IP処理部又はNASサーバの処理性能がネックとなり、レスポンスが低下する。また、TCP/IP処理部又はNAS

サーバに障害が発生した場合、障害が発生したNASの使用を継続できない。

又、上記の問題は、単純に特許文献1のクラスタ構成ストレージシステムにNASを導入するだけでは解決しない。

【0007】

本発明の目的は、NASの使用負荷が上がった場合にも、レスポンスが低下しないNASを提供することにある。

また、本発明の別の目的は、NASに障害が発生した場合にも障害が発生したNASへのアクセスを継続できるNASを提供することにある。

【0008】

【課題を解決するための手段】

本発明の一実施形態はストレージ装置であり、以下の構成を有する。ストレージ装置は、計算機との間で遣り取りされるフレーム（コマンド、データ等含む。パケット等でも良い）の通信プロトコルを処理するインターフェース部（以下「I/F処理部」）と、ファイル操作（ファイルへのデータの書き込み、読み出し、作成、削除等）の処理を行う第一の制御部（以下「NAS部」）と、記憶装置を制御する第二の制御部（以下「キャッシュ制御部」）と、各部を相互に接続する接続網とを有する。本構成において、計算機が送信したフレームを受信したI/F処理部は、受信したフレームをその内容に基づいてNAS処理部に送信したり、キャッシュ制御部に転送したりする。その際I/F処理部は、I/F処理部に割り当てられたアドレスとNAS部との対応関係の情報を有する。

【0009】

又、他の実施形態として、上述のストレージ装置が別のストレージ装置と接続される第二のI/F処理部を有する構成としても良い。尚、この場合、更に、上述の構成からキャッシュ制御部を省略した構成も考えられる。この場合、フレームは、第一のI/F処理部及び第二のI/F処理部を介して他のストレージ装置へ転送される。

【0010】

一方、フレームを受け取ったNAS部は、状態に従い、キャッシュ制御部または、第二のI/F処理部経由で別のストレージ装置にアクセスして所定の処理を行っ

た後、フレームに対する応答をI/F処理部に送信する。その際NAS部は、ファイルと当該ファイルが記憶されているキャッシュ制御部又は他のストレージ装置の対応関係の情報をを用いる。

【0011】

なお、別の実施形態として、I/F処理部とNAS部を複数備える構成とし、I/F処理部とNAS部との対応関係を適宜設定することで、処理負荷の分散を図ったり、障害が発生した時に処理を引き継ぐ構成とすることが考えれる。

また、別の実施形態として、キャッシュ制御部がキャッシュメモリ及びディスク装置を有さない構成としても良い。

【0012】

また、本実施例の好ましい実施形態として、I/F処理部に割り当てられたアドレスとNAS部との対応関係、NAS部が処理するファイルと当該ファイルが記憶されているキャッシュ制御部もしくは他のストレージ装置の対応関係等、各種情報を管理する管理用の計算機（以下、「管理サーバ」）又は同様の管理部をストレージ装置に設ける（又はストレージ装置に接続する）構成がある。管理サーバが、I/F処理部に割り当てられたアドレスとNAS部との対応関係の変更命令をストレージ装置に出すことで、処理負荷の分散を図ったり、障害が発生した時に処理を引き継ぐよう構成変更をすることが可能な構成とする。

【0013】

【発明の実施の形態】

図1は、本発明を適用したシステムの一実施形態の構成を示す図である。システムは、一台以上の計算機（以下「ホスト」）100、一台以上のスイッチ120、ストレージ装置（以下「ストレージ」）130、ストレージ130とスイッチ120経由で接続される一台以上のストレージ（以下「外部ストレージ」）180、管理サーバ110及び管理端末190とを有する。

【0014】

ホスト100、スイッチ120、ストレージ130及び外部ストレージ180は、IPネットワーク175を介して管理サーバ110に接続され、管理サーバ110で動作する図示しない管理ソフトウェアによって統合管理される。なお、

本実施形態では、ストレージ 130 は管理端末 190 を介して管理サーバ 110 に接続する形態をとるものとするが、ストレージ 130 が直接 IP ネットワークに接続される構成であっても良い。

【0015】

ホスト 100 は、CPU 101 やメモリ 102 などをも有する計算機であり、ディスク装置や光磁気ディスク装置などの記憶装置 103 に格納されたオペレーティングシステムやアプリケーションプログラムなどのソフトウェアをメモリ 102 に読み上げ、CPU 101 がメモリ 102 から読み出して実行することで、所定の処理を行う。又、ホスト 100 は、キーボードやマウスなどの入力装置 104 やディスプレイ等の出力装置 105 を具備し、入出力装置 104 でホスト管理者などからの入力を受け付け、出力装置 105 に CPU から指示された情報が出力される。

【0016】

さらにホスト 100 は、ストレージ 130 等外部の装置とネットワーク 176 を介してデータを送受するための一以上のポート 107 と、管理サーバと IP ネットワーク 175 を介してデータ送受するための一以上のインタフェース制御部 106 を有する。

【0017】

管理サーバ 110 は、ホスト 100 と同様の計算機であり、システム全体の運用・保守管理といった所定の処理を行う。CPU 111 によって管理ソフトウェアが実行されると、管理サーバ 110 はインタフェース制御部 116 から IP ネットワーク 175 を介して、システム内の各機器から構成情報、リソース利用率、性能監視情報などを収集する。そして、管理サーバ 110 は収集したそれらの情報を出力装置 115 に出力してストレージの管理者等に提示する。また、管理サーバは、入力装置 114 を介してストレージ管理者からの指示を受信し、受信した運用・保守指示をインタフェース制御部 116 を介して各機器に送信する。

【0018】

スイッチ 120 は、複数のポート 121 及びインタフェース制御部を有する。ポート 121 は、ホスト 100 のポート 107 と IP ネットワーク 176 を介し

て、または、ストレージ130のポート141とIPネットワーク177を介して接続される。又、インタフェース制御部123は、IPネットワーク175を介して管理サーバ110に接続される。スイッチ120は、IPネットワーク内に存在するスイッチが有する経路制御等一般的な処理を行う。

【0019】

ストレージ130はクラスタ構成を有する。具体的には、ストレージ130は複数のプロトコル変換部140、複数のキャッシュ制御部150、1つ以上のNAS部145及び構成管理部160を有し、これらは相互結合網170で相互に接続されている。

【0020】

プロトコル変換部140は、複数のポート141、1つ以上の制御プロセッサ142、メモリ143、及び相互結合網170と接続される転送制御部144を有する。プロトコル変換部140は、ポート141から受信したコマンドを解析し、アクセス対象となるデバイスやNAS部145を特定し、転送制御部144より相互結合網170を介して適当なキャッシュ変換部150やNAS部145等へコマンドやデータを転送する。尚、以下「デバイス」とは、ストレージが提供する物理的又は論理的な記憶領域のことを言う。

【0021】

また、プロトコル変換部140は、ポート141を介して外部ストレージ180など別のストレージと接続されることが可能である。この場合、外部ストレージ180と接続されるプロトコル変換部140は、別のプロトコル変換部140から受信する入出力要求やキャッシュ制御部150から受信する入出力要求を外部ストレージ180へ送信する。これにより、ホスト100又はNAS部145は、外部ストレージ180にデータを書き込んだり（以下「ライト」）、外部ストレージ180からデータを読み出したり（以下「リード」）することができる。

【0022】

なお本実施形態では、ポート141はIPネットワークで主に使用されるTCP/IPと、SCSIを上位プロトコルとしたFiber Channel（以下「FC」）に対応しているとする。NASへのアクセスに使用されるNFS等の通信プロトコルは、TCP/IPプロト

コルの上位プロトコルとして位置する。本実施形態では、記述の簡単のため、NASへのアクセスに使用されるファイルアクセスプロトコルをNFSとするが、CIFS等別のプロトコルでもかまわないし、複数のファイルアクセスプロトコルをサポートしてもかまわない。また、IPネットワークを介してストレージアクセス（デバイスのアクセス）を行う場合に使用されるiSCSI(Internet SCSI)等も、TCP/IPプロトコルの上位プロトコルとして位置し、プロトコル変換部140は、NFS等NASのアクセスに使用されるプロトコルに加え、iSCSI等ストレージが有するデバイスをアクセスするためのプロトコルに対応する。

【0023】

ホスト100がストレージ130又は外部ストレージ180のデバイスをブロックデバイスとして利用したい場合、FC又はIPのネットワーク経由でFCやiSCSI等のプロトコルを用いてストレージ130と通信を行う。

なお、ポート141がFCのプロトコルに対応しない構成でも良い。また、NFS等NASのアクセスに使用されるプロトコルのみに対応する構成でも良い。

【0024】

キャッシュ制御部150は、1つ以上のポート156、各々ポートに接続される1台以上のディスク装置157、1つ以上の制御プロセッサ152、各々の制御プロセッサ152に対応するメモリ153、1つ以上のディスクキャッシュ154、1つ以上の制御メモリ155及び相互結合網170と接続される転送制御部151を有する。

【0025】

制御プロセッサ152は、相互結合網170を介して転送制御部151で受信した、同じキャッシュ制御部150内に存在するディスク装置157への入出力要求を処理する。また制御プロセッサ152は、キャッシュ制御部150がホスト100またはNAS部145に対してディスクアレイのように複数のディスク装置157を1つ又は複数の下位論理デバイスとして提供している場合に、下位論理デバイスとディスク装置157との間の対応関係の管理や、下位論理デバイスに対するアクセス要求をディスク装置157へのアクセス要求に変換する処理などを行う。更に制御プロセッサ152はデータ複製やデータ再配置などの各種処

理を実行する。

【0026】

ディスクキャッシュ154には、ホスト100またはNAS部145からのアクセス要求に対する処理速度を高めるため、ディスク装置157から頻繁に読み出されるデータが予め格納されたり、ホスト100から受信したライトデータが一時的に格納される。尚、ディスクキャッシュ154を用いたライトアフト、即ちホスト100またはNAS部145から受信したライト用のデータ（以下「ライトデータ」）がディスクキャッシュ154に格納された後、ディスク装置157に実際に書き込まれる前にホスト100またはNAS部145に対しライト要求に対する応答を返す場合、ディスクキャッシュ154に格納されているライトデータがディスク装置157に書き込まれる前に消失することを防止するため、ディスクキャッシュ154をバッテリバックアップなどにより不揮発メモリとしたり、媒体障害への耐性向上のため二重化するなど、ディスクキャッシュ154の可用性を向上させておくことが望ましい。

【0027】

制御メモリ155には、キャッシュ制御部150が、ディスク装置157、複数のディスク装置157を組み合わせで構成される物理デバイス又はストレージ130にプロトコル変換部140を介して接続される外部ストレージ180内のデバイス（以下「外部デバイス」）の管理や、外部デバイス又は物理デバイスと下位論理デバイスの対応関係の管理を行うために必要な情報が格納される。なお、制御メモリ155に格納されている制御情報が消失すると、ホスト100やNAS部145はキャッシュ制御部150が有するディスク装置157に格納されているデータへアクセスできなくなるため、制御メモリ155はバッテリバックアップなどにより不揮発メモリとしたり、媒体障害への耐性向上のため二重化するなど、高可用化のための構成を有することが望ましい。

【0028】

また、キャッシュ制御部150の別の構成として、ポート156とディスク157を有さず、外部ストレージ180のデバイスのみを制御する構成としても良い。

【0029】

NAS部145は、一または複数のCPU147、各々の制御プロセッサに対応する一または複数のメモリ149、記憶装置148及び相互結合網170と接続される転送制御部146を有する。記憶装置148に格納された制御プログラムをメモリ149に読み上げ、これをCPU147が実行することで、NAS部145は所定の処理を行う。

【0030】

CPU147は、相互結合網170を介して転送制御部146より受信したコマンドを処理し、コマンドを送信してきたプロトコル変換部140へ応答コマンドを送信する。また、CPU147は、コマンド処理時にデバイスアクセスが必要になった場合（ファイルへのデータの書き込みや読み出し）、そのデバイスを提供（より具体的には、NAS部で実行されているファイルシステムにマウントされるデバイスを提供する）しているキャッシュ制御部150又は外部ストレージ180に対し入出力要求を転送制御部146を介して送信し、その応答を相互結合網170を介して転送制御部146より受信する。尚、デバイスへアクセスを行う場合、NAS部145は、受信したファイルアクセスのコマンドをそのファイルのデータが格納されるデバイスのブロックへのアクセスコマンドに変換する。

【0031】

構成管理部160は、一又は複数の制御プロセッサ162、各々の制御プロセッサに対応する一又は複数のメモリ163、一又は複数の制御メモリ164、記憶装置165、相互結合網170と接続される転送制御部161及びインタフェース制御部166を有する。記憶装置165に格納された制御プログラムをメモリ163に読み上げ、これを制御プロセッサ162が実行することで、ストレージ130の構成管理や障害管理のための所定の処理が実行される。

【0032】

制御プロセッサ162は、インタフェース制御部166を介して接続される管理端末190にストレージ管理者へ提示する構成情報を送信したり、管理者から管理端末190に入力された保守・運用指示を管理端末190から受信して、受信した指示に従い、ストレージ130の構成変更などを行う。ストレージ130

の構成情報は制御メモリ 164 に格納される。制御メモリ 164 に格納された構成情報はプロトコル変換部 140 の制御プロセッサ 142、キャッシュ制御部 150 の制御プロセッサ 152、NAS部 145 の制御プロセッサ 147 によって参照されたり更新されたりすることができる。これにより、ストレージ 130 内のプロトコル変換部 140 及びキャッシュ制御部 150 間で構成情報を共有することができる。

【0033】

なお、構成管理部 160 が障害などにより動作不可に陥った場合、ストレージ 130 全体へのアクセスが不可能になるため、構成管理部 160 内の各構成要素を二重化しておくか、もしくは構成管理部 160 自体をストレージ 130 内に複数搭載して構成管理部 160 自体の二重化をしておくことが望ましい。また、管理端末 190 から個々のキャッシュ制御部 150 への I/F を別途設け、構成管理部 160 が行う制御を個々のキャッシュ制御部 150 と管理端末 190 で分担することで、構成管理部 160 を制御メモリ 164 のみの構成にすることも可能である。さらに、制御メモリ 164 内の情報を個々のキャッシュ制御部 150 の制御メモリ 155 で保持することで、構成管理部 160 を省略することも可能である。

【0034】

相互結合網 170 は、プロトコル変換部 140、キャッシュ制御部 150 及び構成管理部 160 を相互に接続し、これらの各装置間でのデータ、制御情報及び構成情報の送受信を実行する。この相互結合網 170 により、構成管理部 160 がプロトコル変換部 140、キャッシュ制御部 150 および NAS部 145 にストレージ 130 の構成情報を配布したり、プロトコル変換部 140、キャッシュ制御部 150 及び NAS部 145 から構成情報を取得してストレージ 130 の構成を管理したりすることができる。

【0035】

また、相互結合網 170 がプロトコル変換部 140、キャッシュ制御部 150 及び NAS部 145 との間のアクセス要求を転送するので、ホスト 100 はプロトコル変換部 140 の任意のポート 141 から NAS部 145 やキャッシュ制御部 1

50に属する任意の上位論理デバイスへアクセスすることが可能となる。なお、可用性向上の観点から相互結合網170も多重化されていることが望ましい。

【0036】

管理端末190は、CPU192、メモリ193、記憶装置194、構成管理部160と接続するインタフェース制御部191、IPネットワーク175と接続するインタフェース制御部197、ストレージ管理者からの入力を受け付ける入力装置195及びストレージ管理者にストレージ130の構成情報や管理情報を出力するディスプレイ等の出力装置196を有する計算機である。CPU192は記憶装置194に格納されているストレージ管理プログラムをメモリ193に読み出して、これを実行することにより、構成情報の参照、構成変更の指示、特定機能の動作指示などを行い、ストレージ130の保守運用に関して、ストレージ管理者もしくは管理サーバ110とストレージ130間のインタフェースとなる。

【0037】

なお、管理端末190を省略して、ストレージ130を直接管理サーバ110へ接続し、ストレージ130を管理サーバ110で動作する管理ソフトウェアを用いて管理してもよい。

【0038】

外部ストレージ180は、一又は複数のポート181、制御プロセッサ182、メモリ183、ディスクキャッシュ184、一又は複数のディスク装置186及び各々ディスク装置186に接続される一又は複数のポート185を有する。外部ストレージ180は、各々が有するポート181及びスイッチ120を介してストレージ130に接続される。又、外部ストレージ180が有するディスク装置186から構成される論理的な記憶領域である外部デバイスは、ストレージ130によりストレージ130が有する上位論理デバイスとしてNAS部145またはホスト100へ提供される。

【0039】

なお、システムはスイッチ120の他にFCスイッチを有し、FCスイッチ経由で外部ストレージ180がストレージ130と接続されてもよい。また、スイッチを経由せず直接ストレージ130と外部ストレージ180とを接続してもよい。

また、外部ストレージ 180 を設けず使用しないシステム構成としてもよい。

【0040】

制御プロセッサ 182 はメモリ 183 に格納されているプログラムを実行することにより、ポート 181 から受信したディスク装置 186 への入出力要求を処理する。本実施形態では外部ストレージ 180 を、クラスタ構成を有さず、ストレージ 130 より小規模な構成のストレージ装置とするが、ストレージ 130 と同じ構成を有する同規模のストレージ装置であってもかまわない。また、外部ストレージ 180 がディスクキャッシュ 184 や制御プロセッサ 182 の無い JBOD (Just Bunch Of Disk) のような単体ディスク装置グループであっても構わない。

【0041】

以下、本実施形態のシステムの動作概要について簡単に説明する。

ストレージ 130 の有するディスク装置 157 または外部デバイスからなる上位論理デバイスの一部もしくは全ては、ファイルアクセス処理を行う NAS 部 145 に提供される。

【0042】

ホスト 100 がストレージ 130 に記憶されるファイルを使用するためにストレージ 130 に送信したコマンド（ファイルアクセスのコマンド）は、ストレージ 130 のポート 141 の内、ファイルアクセス用（NFS）のアドレスが割り当てられているポート 141 に到着する。そのポート 141 を有するプロトコル変換部 140 は、そのコマンドを解析し、そのコマンドを処理すべき NAS 部 145 を選択し、そのコマンドを相互結合網 170 経由で選択した NAS 部 145 に送信する。

【0043】

コマンドを受信した NAS 部 145 は、そのコマンドで指定されるファイルに対する処理を行う。コマンドの処理中にデバイス（具体的にはそのデバイスに格納されているファイル）へのアクセスが必要となった場合、そのデバイスがキャッシュ制御部 150 が提供する上位論理デバイスであれば、NAS 部 145 は当該キャッシュ制御部 150 へアクセスコマンドを送信し、必要なデータを得たり、デ

ータをキャッシュ制御部 150 に格納したりする。尚、上位論理デバイスに外部デバイスが含まれる場合、キャッシュ制御部 145 は、プロトコル変換部 145 を介してコマンドを外部ストレージ 180 へ送信する。また、NAS 部 145 が外部ストレージ 180 が提供するデバイスを直接利用している場合、NAS 部 145 は、当該外部ストレージが接続されているプロトコル変換部 140 にアクセスコマンドを送信し、必要なデータを得たり、データを外部ストレージ 180 が提供するデバイスに格納する。

【0044】

コマンドの処理を終了した NAS 部 145 は、応答コマンドを作成し、相互結合網 170 を経由して先に受信したコマンドの送信元のプロトコル変換部 140 に応答コマンドを送信する。応答コマンドを受信したプロトコル変換部 140 は、応答コマンドをホスト 100 へ送信する。

【0045】

なお、以降の説明では、表現の簡略化のため、プロトコル変換部 140 を PA、キャッシュ制御部 150 を CA、構成管理部 160 を MA、管理端末 190 を ST、NAS 部 145 を AA と表記する。

【0046】

また、本発明の実施形態では、ストレージ 130 内の各ディスク装置、物理デバイス、下位又は上位論理デバイスの関係（以下「デバイス階層」）を次の通りとする。

まず、CA 内でディスク装置 157 複数台によるディスクアレイが構成され、このディスクアレイは CA によって物理デバイスとして管理される。また、PA に接続された外部ストレージ 180 が提供する外部デバイスは、PA で認識された後、MA にて管理される。

【0047】

更に CA では、CA 内に存在する物理デバイス又は当該 CA を介してアクセスされる外部デバイスに対して下位論理デバイスを割り当てる。下位論理デバイスは各 CA 内において管理される論理デバイスであり、その番号は各 CA 毎に独立に管理される。下位論理デバイスは、MA によって上位論理デバイスと対応付け

られ、ストレージ 130 が提供する論理的な記憶装置（デバイス）としてホスト 100 または AA に提供される。

【0048】

即ち、ホスト 100 または AA が認識するのはストレージ 130 の上位論理デバイスであり、ホスト 100 は上位論理デバイスを識別するための情報を用いて、ストレージ 130 若しくは外部ストレージ 180 に格納されているデータにアクセスする。ただし、AA は、管理情報に従い PA を介して直接外部ストレージ 180 の外部デバイスにアクセスすることもできる。この場合、外部デバイスは特定の CA には割り当てられていない。

【0049】

尚、個々のディスク装置 157 を 1 つの物理デバイスおよび 1 つの下位論理デバイス及び上位論理デバイスとしてもよい。また、複数の物理デバイスを 1 つ又は複数の下位論理デバイスに対応付けたり、複数の下位論理デバイスに 1 つ又は複数の上位論理デバイスを対応付けても良い。

以下、本システムの動作を説明する前に、本システムの各構成要素が有する情報について説明する。

【0050】

図 2 は、ストレージ 130 の各構成要素の制御メモリおよびメモリに格納される、制御情報、管理情報とストレージ制御処理のためのプログラムの一例を示したソフトウェア構成図である。

【0051】

ストレージ 130 のデバイス階層を管理するための管理情報として、CA の制御メモリ 155 に下位論理デバイス管理情報 201、物理デバイス管理情報 202 及びキャッシュ管理情報 203 が、MA の制御メモリ 164 に上位論理デバイス管理情報 204、外部デバイス管理情報 205 及び LU パス管理情報 206 が格納されている。

【0052】

CA 内の制御メモリ 155 および MA に格納されている各制御情報は、各 CA、各 PA、各 AA、及び MA 内の各制御プロセッサから参照・更新可能であるが

、その際に相互接続網 1 7 0 を介したアクセスが必要となる。よって、処理性能向上のため、各制御プロセッサで実行される処理に必要な制御情報の複製が、各部位（即ち、CA、PA、AA及びMA）内のメモリに保持される。

【 0 0 5 3 】

各部位は、当該部位が管理する制御情報が構成変更により更新された場合は、相互接続網 1 7 0 を介してその旨を他の部位に通知し、最新情報を当該部位の制御メモリから他の部位のメモリへ取り込ませる。なお、制御情報が更新された場合に更新を他の部位に知らせる別の方法として、例えば、各部位は、自部位内の制御メモリに保持した構成情報毎に制御メモリ上に更新有無を示すフラグを設け、各部位の制御プロセッサは処理開始時もしくは各構成情報を参照する毎に同フラグを参照して更新有無をチェックする方法などが考えられる。

【 0 0 5 4 】

なお、各部位のメモリには上述の制御情報の複製に加えて、各部位内の制御プロセッサで動作する制御プログラムが格納される。

【 0 0 5 5 】

NASの処理（ファイルアクセス処理、障害時のフェイルオーバー等）にかかわる管理情報としては、MAの制御メモリ 1 6 4 にNAS管理情報 2 2 3、ポート管理情報 2 2 4 が格納されている。

【 0 0 5 6 】

PAのメモリ 1 4 3 には、上位論理デバイス管理情報 2 0 4 の全て又は当該PAの処理にかかわる情報の複製 2 1 4、外部デバイス管理情報 2 0 5 の全て又は当該PAの処理にかかわる情報の複製 2 1 5、LUパス管理情報 2 0 6 の全て又は当該PAの処理にかかわる情報の複製 2 1 6、ポート管理情報 2 2 4 の全て又は当該PAの処理にかかわるポート管理情報の複製 2 2 2、NAS管理情報 2 2 3 の全て又は当該PAの処理にかかわるNAS管理情報の複製 2 2 1、ポート 1 4 1 または転送制御部 1 4 4 から受信したデータもしくはコマンドを振り分ける要求振り分け処理プログラム 2 5 1、LUパス定義処理プログラム 2 5 2、外部デバイス定義処理プログラム 2 5 3、NAS／ポート障害処理プログラム 2 3 1 及びファイルアクセス処理プログラム 2 2 9 が格納されている。

【 0 0 5 7 】

CAの制御メモリ155には、下位論理デバイス管理情報201、キャッシュ管理情報203、物理デバイス管理情報202が格納される。メモリ153には、下位論理デバイス管理情報201の複製211、物理デバイス管理情報202の複製212、外部デバイス管理情報205の全て又は当該CAの処理にかかわる情報の複製215、PAまたはAAから受信した入出力コマンドを処理するコマンド処理プログラム254、論理デバイス定義処理プログラム255、ライトアフタ処理プログラム及び外部デバイス接続変更処理プログラム256が格納されている。

【 0 0 5 8 】

MAの制御メモリ164には、上位論理デバイス管理情報204、外部デバイス管理情報205、LUパス管理情報206、NAS管理情報223及びポート管理情報224が格納される。メモリ163には、全デバイス管理情報(即ち、上位論理デバイス管理情報204、下位論理デバイス管理情報201、物理デバイス管理情報202及び外部デバイス管理情報205)の複製210、論理デバイス定義処理プログラム255、外部デバイス接続変更処理256、LUパス定義処理252、NAS/ポート障害処理プログラム231、外部デバイス定義処理プログラム253、NAS管理情報変更処理プログラム225及びポート管理情報変更処理プログラム226が格納されている。

【 0 0 5 9 】

AAのメモリ149には、NAS管理情報223の全て又は当該AAの処理にかかわる情報の複製227、外部デバイス管理情報205の全て又は当該AAの処理にかかわる情報の複製215、当該AAのキャッシュ管理情報、NAS/ポート障害処理プログラム231及びファイルアクセス処理プログラム229が格納されている。

【 0 0 6 0 】

なお、本実施形態では、全ての管理情報(下位論理デバイス管理情報201、物理デバイス管理情報202、上位論理デバイス管理情報204、外部デバイス管理情報205、LUパス管理情報206、NAS管理情報223、ポート管理情報

2 2 4 及び各複製) は、ストレージ 1 3 0 の工場出荷時にあらかじめ、初期値が設定されているものとする。その後、適宜、ストレージ管理者または、管理サーバからの制御により後述の方法を用いて値が設定変更され、運用される。

【0 0 6 1】

図 1 5 は、S T の制御メモリに格納される、制御情報とストレージ制御処理のためのプログラムの一例を示したソフトウェア構成図である。メモリ 1 9 3 には、全デバイス管理情報(即ち、上位論理デバイス管理情報 2 0 4、下位論理デバイス管理情報 2 0 1、物理デバイス管理情報 2 0 2 及び外部デバイス管理情報 2 0 5)の複製 2 1 0、NAS管理情報 2 2 3 の複製 2 3 1、ポート管理情報 2 2 4 の複製 2 3 2、論理デバイス定義処理プログラム 2 5 5、LUパス定義処理プログラム 2 5 2、外部デバイス定義処理プログラム 2 5 3、NAS管理情報変更処理プログラム 2 2 5、ポート管理情報変更処理プログラム 2 2 6 及びNAS／ポート障害処理プログラム 2 3 1 が格納される。

【0 0 6 2】

図 3 は、上位論理デバイス管理情報 2 0 3 の一構成例を示す図である。上位論理デバイス管理情報は、各上位論理デバイス毎に、上位論理デバイス番号から対応外部デバイス番号までの情報の組を保持するエントリ 3 1 ~ 3 9 を有する。なお、切替進捗ポインタを使用しない場合はエントリ 3 9 を有しなくても良い。

【0 0 6 3】

エントリ 3 1 には、対応する上位論理デバイスを識別するために上位論理デバイスに M A が割り当てた番号が格納される。エントリ 3 2 には、上位論理デバイス番号により特定される上位論理デバイスの記憶容量が格納される。エントリ 3 3 には、当該上位論理デバイスに対応付けられている下位論理デバイスの番号と当該下位論理デバイスが属する C A の番号が格納される。上位論理デバイスが未定義の場合、エントリ 3 3 には無効値が設定される。尚、この下位論理デバイス番号が、当該下位論理デバイスを管理する C A が保持する下位論理デバイス管理情報 2 0 1 のエントリ番号となる。

【0 0 6 4】

エントリ 3 4 には、対応する上位論理デバイスの状態を示す情報が設定される

。状態としては、「オンライン」、「オフライン」、「未実装」及び「障害オフライン」が存在する。「オンライン」は、当該上位論理デバイスが正常に稼動し、ホスト100またはAAからアクセスできる状態であることを示す。「オフライン」は、当該上位論理デバイスは定義され、正常に稼動しているが、LUパスが未定義であるなどの理由で、ホスト100またはAAからはアクセスできない状態にあることを示す。「未実装」は、当該上位論理デバイスが定義されておらず、ホスト100またはAAからアクセスできない状態にあることを示す。「障害オフライン」は、当該上位論理デバイスに障害が発生してホスト100またはAAからアクセスできないことを示す。

【0065】

なお、本実施形態では簡単のため、ストレージ130の工場出荷時にあらかじめ、ディスク装置157で作成された物理デバイスへ下位論理デバイスが割り当てられ、更に下位論理デバイスに上位論理デバイスが割り当てられているものとする。このため、下位論理デバイスに割当済みの上位論理デバイスについてのエントリ34の初期値は「オフライン」、その他の上位論理デバイスについては上位論理デバイスが定義されていないことになるので、エントリ34の初期値は「未実装」となる。

【0066】

エントリ35には、ポート番号、ターゲットID及びLUNが登録される。ここで登録されるポート番号は、対応する上位論理デバイスが複数のポート141またはAAの転送制御部146のうちどのポートまたはAAの転送制御部146に接続されているかを表す情報、即ち対応する上位論理デバイスにアクセスするために用いられるポート141またはAAの転送制御部146の識別情報である。ここでポート141又はAAの転送制御部146の識別情報は、各ポート141及びAAの転送制御部146に割り振られているストレージ130内で一意な番号であり、エントリ35には、当該上位論理デバイスがLUN定義されているポート141又はAAの転送制御部146の番号が記録される。

【0067】

また、ターゲットIDとLUNは、当該上位論理デバイスを識別するための識

別子である。本実施形態においては、上位論理デバイスを識別するための識別子として、SCSIプロトコルでホスト100またはAAからデバイスをアクセスする場合に用いられるSCSI-ID及び論理ユニット番号(LUN)を用いる。

【0068】

エントリ36には、当該上位論理デバイスへのアクセスが許可されているホスト100またはAAを識別するホスト名が登録される。ホスト名としては、ホスト100のポート107に付与されたWWN(World Wide Name)やiSCSIネームなど、ホスト100又はポート107及びAAを一意に識別可能な値であれば何を用いてもよい。ストレージ130は、このほかに、各ポート141のWWNやMAC(Media Access Control)アドレス、IPアドレスなどの属性に関する管理情報を保持する。

【0069】

エントリ37には、当該上位論理デバイスへの入出力要求の処理形態を示す「CA経由」、「PA直結」又は「PA直結移行中」という値が格納される。

「CA経由」は、当該上位論理デバイスが外部デバイスと対応付けられていない場合や、当該上位論理デバイスが外部デバイスに対応付けられていて、かつ当該上位論理デバイスに対する入出力要求に対してCA150での入出力処理が必要な処理形態を示す。

【0070】

「PA直結」は、当該上位論理デバイスが外部デバイスに対応付けられており、かつ、CAを経由せずに直接PA間で当該上位論理デバイスへの入出力要求を転送できる処理形態を示す。

「PA直結移行中」は、外部デバイス接続形態を「CA経由」から「PA直結」へ切り替える過渡状態、すなわち、後述する切替処理が対応CAにて実行中であることを示す。

【0071】

エントリ38には、当該上位論理デバイスが外部デバイスに対応付けられている場合にはその外部デバイスの識別番号が、そうでない場合は無効値が設定され

る。

【0072】

エントリ 39 には、切替進捗ポインタが登録される。切替進捗ポインタは、当該上位論理デバイスが「PA直結移行中」である場合に使用され、当該上位論理デバイスが有する記憶領域のうち、外部接続の接続形態変更処理が未完了な領域の先頭アドレスを示す情報である。なお、「PA直結移行中」処理を行わない場合には、エントリ 39 を省いてもかまわない。

【0073】

図 4 は、LUパス管理情報 206 の一構成例を示す図である。LUパス管理情報 206 は、PAが有する各ポート 141 又は AAの転送制御部 146 に対して定義されている有効な LUN毎に、情報を保持する為のエントリ 41～43 を有する。エントリ 41 には、ポート 141 又は AAの転送制御部 146 に定義された（割り当てられた）LUNのアドレスが格納される。

【0074】

エントリ 42 には、当該 LUNが割り当てられている上位論理デバイスの番号が格納される。エントリ 43 には、当該 LUNに対してアクセスを許可されているホスト 100 又は AAを示す情報、具体的には WWNや転送制御部 146 の番号等が登録される。

【0075】

尚、一つの上位論理デバイスに対して複数のポート 141 又は AAの転送制御部 146 が定義され（割り当てられ）ており、複数のポート 141 又は AAから当該上位論理デバイスにアクセスできる場合がある。この場合、当該複数のポート 141 又は AAの LUN各々に関する LUパス管理情報 206 のエントリ 43 に登録された値の和集合が、当該上位論理デバイスに関する上位論理デバイス管理情報 203 のエントリ 36 に保持される。

【0076】

図 5 は、下位論理デバイス管理情報 201 の一構成例を示す図である。各 CAは、自 CAに属する各下位論理デバイス毎に、下位論理デバイス番号から対応上位論理デバイス番号までの情報の組を登録するエントリ 51～56 を有する。な

お、切替進捗ポインタが使用されない場合は、エントリ 5 6 は省いてもかまわない。

【0 0 7 7】

エントリ 5 1 には、対応する下位論理デバイスを識別するための識別番号が登録される。エントリ 5 2 には、エントリ 5 1 により特定される下位論理デバイスの記憶容量が格納される。

【0 0 7 8】

エントリ 5 3 には、当該下位論理デバイスに対応付けられる、C A の物理デバイスの識別番号又は外部ストレージ 1 8 0 に存在する外部デバイスの識別番号が格納される。当該下位論理デバイスに対して物理デバイス又は外部デバイスが割り当てられていない場合、エントリ 5 3 には無効値が設定される。なお、エントリ 5 3 に登録されるデバイス番号は、当該下位論理デバイスを管理している C A が保持する物理デバイス管理情報 2 0 2、もしくは M A が保持する外部デバイス管理情報 2 0 5 のエントリ番号と一致する。

【0 0 7 9】

エントリ 5 4 には、当該下位論理デバイスの状態を示す情報が設定される。設定される情報の内容は、上位論理デバイス管理情報 2 0 3 のエントリ 3 4 と同じであるため、説明は省略する。エントリ 5 5 には、当該下位論理デバイスに対応付けられている上位論理デバイス番号が設定される。

【0 0 8 0】

エントリ 5 6 には、切替進捗ポインタが登録される。このポインタは上位論理デバイス管理情報 2 0 4 の切替進捗ポインタと同じで、当該下位論理デバイスに対応する上位論理デバイスが「P A 直結移行中」である場合に使用され、移行中の上位論理デバイスに対応する下位論理デバイスのうち、外部接続の接続形態変更処理が未完了な記憶領域の先頭アドレスを示す情報である。

【0 0 8 1】

図 6 は、C A のディスク装置 1 5 7 から構成される物理デバイスを管理するための物理デバイス管理情報 2 0 2 の一構成例を示す図である。各 C A は、自 C A が有する物理デバイス毎に、物理デバイス番号からディスク内サイズの情報の組

を保持する為のエントリ 61～69 を有する。

【0082】

エントリ 61 には、物理デバイスを識別するための識別番号が登録される。エントリ 62 には、エントリ 61 に登録された値で特定される物理デバイスの記憶容量が格納される。エントリ 63 には、当該物理デバイスに対応付けられる下位論理デバイスの番号が格納される。当該物理デバイスが下位論理デバイスへ割り当てられていない場合、エントリ 63 には無効値が設定される。

【0083】

エントリ 64 には、当該物理デバイスの状態を示す情報が設定される。状態としては、「オンライン」、「オフライン」、「未実装」及び「障害オフライン」が存在する。「オンライン」は、当該物理デバイスが正常に稼動し、下位論理デバイスに割り当てられている状態であることを示す。「オフライン」は、当該物理デバイスは定義され正常に稼動しているが、下位論理デバイスに未割り当てであることを示す。「未実装」は、当該物理デバイスがディスク装置 157 に対応付けられていない状態にあることを示す。「障害オフライン」は、当該物理デバイスに障害が発生して下位論理デバイスに割り当てられないことを示す。

【0084】

なお、本実施形態では、簡単のため、物理デバイスは製品の工場出荷時にあらかじめディスク装置 157 に対応付けられて作成され、かつ下位論理デバイスにも割り当てられているものとする。このため、利用可能な物理デバイスについてはエントリ 64 の初期値は「オンライン」状態、その他は「未実装」状態となる。

【0085】

エントリ 65 には、当該物理デバイスが割り当てられたディスク装置 157 の R A I D レベル、データディスクとパリティディスク数など R A I D 構成に関連する情報が保持される。同じように、エントリ 66 には、R A I D におけるデータ分割単位(ストライプ)長が保持される。エントリ 67 には、当該物理デバイスが割り当てられた R A I D を構成する複数のディスク装置 157 各々の識別番号が保持される。このディスク装置 157 の識別番号は、C A 内でディスク装置 1

57を識別するために付与された一意な値である。

【0086】

エントリ68とエントリ69には、当該物理デバイスが各ディスク装置157内のどの領域に割り当てられているかを示す情報（開始オフセット及びサイズ）が登録される。本実施例では簡単のため、全物理デバイスについて、RAIDを構成する各ディスク装置157内のオフセットとサイズを統一している。

【0087】

図7は、ストレージ130に接続された外部ストレージ180の外部デバイスを管理するための、外部デバイス管理情報205の一構成例を示す図である。ストレージ130のMAは、外部ストレージ180が有する外部デバイス毎に、外部デバイス番号からターゲットポートID／ターゲットID／LUNリストまでの情報の組を保持する為のエントリ71～79を有する。

【0088】

エントリ71には、ストレージ130のMAが当該外部デバイスに対して割り当てた、ストレージ130内で一意な値が格納される。エントリ72には、エントリ71に登録された値より特定される外部デバイスの記憶容量が格納される。

【0089】

エントリ73には、当該外部デバイスに対応付けられているストレージ130内の上位論理デバイスの番号が登録される。エントリ74には、当該外部デバイスに対応付けられ、当該外部デバイスに対するホスト100からのアクセス要求を処理する、ストレージ130のCAの番号と、当該外部デバイスに対応付けられている下位論理デバイスの番号が格納される。また、AAが直接当該外部デバイスを使用する場合は、CAの番号ではなく、AAの番号が格納される。なお、当該外部デバイスに対して上位論理デバイスもしくは下位論理デバイスが割り当てられていない、もしくはAAが直接使用しない場合、エントリ73もしくはエントリ74には無効値が設定される。

【0090】

エントリ75には、当該外部デバイスの状態を示す情報が設定される。各状態の種類は物理デバイス管理情報202のエントリ64と同じである。ストレージ

1 3 0 は初期状態では外部ストレージ 1 8 0 を接続していないため、エントリ 7 5 の初期値は「未実装」となる。エントリ 7 6 には、当該外部デバイスを有する外部ストレージ 1 8 0 の識別情報が登録される。外部ストレージ 1 8 0 の識別情報としては、外部ストレージ 1 8 0 のベンダ識別情報と各ベンダが一意に割り振る製造シリアル番号の組み合わせなどが考えられる。

【0 0 9 1】

エントリ 7 7 には、当該外部デバイスに対して当該外部デバイスを有する外部ストレージ 1 8 0 が割り振ったデバイスの識別番号が格納される。なお、外部デバイスは外部ストレージ 1 8 0 の論理デバイスであるから、エントリ 7 7 には外部ストレージ 1 8 0 における論理デバイス番号が格納される。

【0 0 9 2】

エントリ 7 8 には、当該外部デバイスへアクセス可能なストレージ 1 3 0 のポート 1 4 1 に割り振られた識別番号と、当該ポート 1 4 1 が属する P A の識別番号が登録される。複数のポート 1 4 1 から当該外部デバイスへアクセスできる場合には、複数のポート識別番号と複数の P A 識別番号が登録される。

【0 0 9 3】

エントリ 7 9 には、当該外部デバイスが外部ストレージ 1 8 0 の 1 つ以上のポート 1 8 1 に L U N 定義されている場合、それらのポート 1 8 1 のポート I D および当該外部デバイスに割り当てられたターゲット I D / L U N が 1 つ又は複数個保持される。なお、ストレージ 1 3 0 の P A が外部デバイスにアクセスする場合（P A から外部デバイスに対する入出力要求を送信する場合）には、当該外部デバイスが属する外部ストレージ 1 8 0 によって当該外部デバイスに割り当てられたターゲット I D 及び L U N が、当該外部デバイスを識別するための情報として用いられる。

【0 0 9 4】

図 1 6 は、NAS 管理情報 2 2 3 の一構成例を示す図である。NAS 構成情報 2 2 3 は、各 A A ごとに、A A 番号～代替 A A 番号の情報の組を登録するエントリ 1 6 0 1 ～1 6 0 6 を有する。

エントリ 1 6 0 1 には、ストレージ 1 3 0 内で A A を識別するための識別情報

が登録される。

【0095】

エントリ 1602 には、ホスト 100 がファイルアクセス用の通信コマンドを送信する際に送信先を指定するための IP アドレス、当該 AA が管理するファイルが存在するディレクトリ名、当該ディレクトリ名で指定されるディレクトリの情報とディレクトリ下に存在するファイルを格納するために割り当てられている上位論理デバイスの番号もしくは外部デバイス番号、及び当該上位論理デバイスを AA に対して提供している CA の番号の組からなる情報のリストが登録される。なお、IP アドレスのみを登録し、登録された IP アドレス先に来た全要求を当該 AA が処理するようにしても良い。

【0096】

エントリ 1602 に複数の IP アドレスが登録されている場合、ホスト 100 からは、複数の NAS がシステム内にあるかのように見える。

また、1 つの IP アドレスに複数の AA を割り当てて、例えばディレクトリ毎に異なる AA が処理するようにすることもエントリ 1602 の設定で可能である。また、AA が複数の上位論理デバイスを例えばディレクトリ毎に使い分ける場合もエントリ 1602 の設定で可能である。

【0097】

エントリ 1603 には、AA の処理状態を示す情報が登録される。登録される情報には、「オフライン」、「障害」及び「オン中」の 3 つが有る。

「オフライン」とは、AA が初期設定処理中だったり、他 AA からの情報引継ぎ処理中であってホスト 100 からの要求を受け付けない状態を示す。

「障害」とは、AA に障害が発生した状態を示す。

「オン中」とは、AA がホスト 100 からの要求を受け付け可能な状態を示す。

【0098】

エントリ 1604 には、AA のアクセスモードを示す情報が登録される。アクセスモードを示す情報とは AA が他の部位（CA や PA）と連携して動作を行うかどうかを示す情報であり、「連携」と「非連携」の二種類がある。

「非連携」が設定される時には、A Aの処理にかかわるコマンドおよびデータは、全てA Aを経由する。具体的には、ホスト100からP Aが受信したコマンドはA Aに送信され、A Aは、必要に応じ入出力要求をC AもしくはP A経由で外部ストレージ180に送信する。又、その応答コマンドおよびデータはC AもしくはP AからA Aに送信され、処理終了後、A Aは応答コマンドをホスト100からのコマンドを受信したP Aに送信する。

【0099】

「連携」が設定される時には、A Aの入出力要求に基づく、C A又はP A経由で接続された外部ストレージ180から送信されるデータはA Aを介さず直接ホスト100からコマンドを受信したP Aへ送信され、C A又はP A経由で接続された外部ストレージ180へ送信するデータは、ホスト100からコマンドを受信したP Aから直接C A又は外部ストレージ180と接続されているP Aに対し送信される。これにより、大量のデータがA Aを経由しないため、A Aの処理負荷と相互接続網170の負荷が軽減される。なお、本実施形態では、簡単のため、ストレージ130の工場出荷時にあらかじめ、初期値として「非連携」がエントリ1604に設定されており、その後、適宜、ストレージ管理者または、管理サーバからの制御により後述の手段を用いて値が設定変更されるものとする。「連携」は、例えばサイズの大きいファイルに対するアクセスが多いと推測される環境において、処理負荷の軽減のためにストレージ管理者により採用される。なお、エントリ1604を省略し、A Aとして「連携」もしくは「非連携」のいずれか一方を提供することにしてもよい。

【0100】

エントリ1606には、当該A Aに障害が発生した場合に、当該A Aの管理情報を引き継ぎ処理を行う代替A Aの番号が設定される。代替A Aが無い場合には、無効値が設定される。代替A Aを設けない場合には、エントリ1606は省略されてもよい。

【0101】

図17は、ポート管理情報224の一構成例を示す図である。ポート管理情報224は、ポート141がIPネットワークと接続される場合にその接続状態を管

理する際にMA等の各部位で用いられる。尚、ポート141がFCネットワークに接続される場合には、そのポート141は、ポート管理情報224により管理されない。ポート管理情報224は、IPネットワークと接続されるポート141ごとに、そのポートに関する情報を登録するエントリ1701～1708を有する。

【0102】

エントリ1701には、ポート141をストレージ130内で一意に識別するために割り付けられた番号が設定される。エントリ1702には、当該ポート141のMACアドレスが登録される。エントリ1703には、当該ポート141に割り当てられたIPアドレスのリストが登録される。エントリ1704には、当該ポート141が処理できるプロトコル（例えば、iSCSI、NFS等）のリストが登録される。ポート141は、自ポートのエントリに登録されたMACアドレス1702、IPアドレス1703、プロトコル1704に従い通信処理を行う。なお、処理できないプロトコルのデータをポート141が受信した場合には、当該データは、ポート141にて廃棄処理される。

【0103】

エントリ1705には、当該ポート141の状態が登録される。状態には「通信可能」、「ネットワーク非接続」、「通信不可能」及び「障害」の四種類がある。

「通信可能」は、ポート141に障害が無くかつネットワークに接続されており、ホスト100との通信が可能な状態にあることを示す。「ネットワーク非接続」は、ポート141にケーブル等が接続されておらずネットワークが接続されていない状態を示す。「通信不可能」は、ポート141がネットワークに接続されているが、ネットワークの問題でホスト100と通信ができない状態にあることを示す。「障害」は、ポート141に障害ある状態を示す。

【0104】

エントリ1707には、ポート141が外部ストレージ180に接続する場合には、当該ポート141に接続された外部ストレージ180の外部デバイス番号が登録される。ポート141が外部ストレージ180に接続されていない場合には

、無効値が設定される。

【0105】

エントリ 1708 には、当該ポート 141 に障害が発生した場合に、当該ポート 141 に割り当てられた IP アドレスのリスト 1703 を引継ぎ当該 IP アドレスに対するアクセスを続行するポート 141（以下「代替ポート」）のポート番号が登録される。なお、代替ポートが通信に使用されていなかった場合に引継ぎ処理を行う場合、IP アドレスのリストだけでなく、MAC アドレスも代替ポートへ引き継いでもかまわない。

【0106】

なお、ポート管理情報 224 で FC プロトコルに対応するするポートも管理する場合、前述の情報に加え、エントリ 1702 には WWN が、エントリ 1703 にはポート ID が、エントリ 1704 には FC を示す情報が登録される。

【0107】

以下、ホスト 100 がストレージ 130 に対してファイルアクセスを行った場合のストレージ 130 内部での処理手順について説明する。ファイルアクセスの処理は、具体的には、要求振り分け処理、ファイルアクセス処理、連携モード処理、コマンド処理及びライトアプタ処理の各処理から構成される。以下、順に説明する。

【0108】

図 11 は、要求振り分け処理 251 の処理フローの一例を示す図である。要求振り分け処理 251 は、PA のポート 141 においてホスト 100 または外部ストレージ 180 から受け取った入出力要求やデータ、および転送制御部 144 において CA、AA 又は別の PA から受け取った入出力要求やデータやコマンドを、CA、AA、他の PA、ホスト 100 又は外部ストレージ 180 等、適切な対象部位へ PA が振り分ける処理である。

【0109】

ポート 141 又は転送制御部 144 からデータやコマンドなどを含んだフレームを受信した PA は（ステップ 1101）、フレーム内のプロトコルの種別および当該プロトコルのコマンドの種別によって処理を切り分ける。なお、プロトコ

ルの種別は、当該ポート 1 4 1 がサポートするネットワーク（IPもしくはFC）と受信したフレーム内のデータから判別することができる（ステップ 1 1 0 2）。

【0 1 1 0】

当該フレームが F C P コマンドフレームの場合もしくは iSCSI コマンドであった場合、P A は、当該フレームの送信元によって処理を切り分ける。なお、本実施形態では、コマンドフレームは、ポート 1 4 1 を介してホスト 1 0 0 から受信する場合と、転送制御部 1 4 4 を介して外部ストレージ 1 8 0 をアクセスする C A、A A 又は他 P A から受信する場合の 2 つがある（ステップ 1 1 0 3）。

【0 1 1 1】

まず、送信元がホスト 1 0 0 の場合、P A は、受信したコマンドのヘッダ情報に含まれるアクセス対象の上位論理デバイスを表す LUN に基づいて、L U パス管理情報の複製 2 1 6 及び上位論理デバイス管理情報の複製 2 1 4 を参照し、当該上位論理デバイスが外部デバイスに対応付けられており、かつデバイスアクセスモードが「C A 経由」でないかを判断する。当該上位論理デバイスが外部デバイスに対応付けられておりかつデバイスアクセスモードが「P A 直結」である場合、P A は更に、当該コマンドの要求種別が C A での処理不要であるかを判断する。これらの条件を全て満たす場合、P A は、当該要求のアクセス形態は P A 直結であると判断し、いずれかの条件を満たさない場合は当該要求のアクセス形態は C A 経由であると判断する（ステップ 1 1 0 4 ～ 1 1 0 5）。

【0 1 1 2】

ステップ 1 1 0 5 でアクセスモードが「C A 経由」でなく、コマンド要求種別が C A 処理不要の場合、P A はさらにアクセスモードが「P A 直結移行中」でかつ当該コマンド内 L B A とブロック数で示されるアクセス領域が当該上位論理デバイスの切替進捗ポイントよりも後にあるかを判断する（ステップ 1 1 1 9）。アクセスモードが「P A 直結移行中」で無い又は P A が受信したコマンドが進捗ポイントよりも前、すなわち P A 直結へ切替済みの領域を対象としている場合には、P A は、コマンドのアクセス先を外部デバイスを有する外部ストレージ 1 8 0 が接続された P A とする（ステップ 1 1 0 6）。

【0 1 1 3】

ステップ1119でアクセスモードが「PA直結移行中」でかつアクセス先が未切替な部分であると判断された場合及びステップ1105でCA経由アクセスであると判断した場合、PAは上位論理デバイス管理情報の複製214を参照して、受信したコマンドのアクセス対象となっている上位論理デバイスが対応するCA番号などを算出し、これを入出力処理の要求先、当該フレームの送信元を入出力処理の要求元として、当該コマンドのルーティング制御情報として記憶する（ステップ1107）。

【0114】

なお、PA直結移行中処理を行わない場合、ステップ1119は省略してもかまわない。この場合、図3のエントリ39は省略される。

【0115】

一方、ステップ1103の切り分け処理で、当該フレームの送信元がホスト100以外、すなわちCA、AA又は他PA140の場合、PAは当該フレームのヘッダ情報で指定されている外部ストレージ180の外部デバイスを転送先に決定し、これを当該入出力処理の要求先、当該フレームの送信元を入出力処理の要求元として当該コマンドのルーティング制御情報を登録する（ステップ1108、1109）。

【0116】

一方、ステップ1102の切り分け処理で、PA140で受信したフレームの種別がデータフレーム、受信準備完了通知フレーム又は完了報告フレームの場合、PAは、当該フレームに対応する入出力処理のルーティング制御情報を参照し、当該フレームの送信元部位の交信相手先を当該フレームの転送先として決定する。すなわち、PAが当該フレームを外部ストレージ180から受信した場合は、当該入出力処理の要求元であるホスト100が接続されたPA、CA又はAAがフレームの転送先となる。又、PAが当該フレームをホスト100から受信した場合は、当該入出力処理の要求先であるCA又は外部ストレージ180が接続されたPAが当該フレームの転送先として決定される。さらに、PAが当該フレームをCA又は他PAから受信した場合は、当該入出力処理の交信相手先（即ち当該フレームの送信元が当該入出力処理の要求元ならば要求先）であるホスト1

00又は外部ストレージ180が、それぞれ当該フレームの転送先として決定される（ステップ1110）。

【0117】

ステップ1106、1107、1109及び1110で受信フレームの転送先が決まったら、PAは転送先部位によって処理を切り分ける（ステップ1111）。

転送先が他のPAの場合、PAは、当該入出力要求のルーティング制御情報に登録されている転送先のアドレスを当該フレームの送信先アドレスへ設定する。すなわち、送信先が外部デバイスの場合、PAは、受信したフレームのデスティネーションIDをルーティング制御情報に当該入出力処理の要求先として登録されている外部デバイスのターゲットポートID（即ちポート181のポートID）へ、また、同フレーム内のLUNなどのアドレス情報を当該外部デバイスのアドレス（即ち外部デバイス管理情報を参照して取得したLUN等）に書き換える。送信先がホスト100の場合、PAは、受信フレームのデスティネーションIDをルーティング制御情報に当該入出力処理の要求元として登録されているホスト100のポート107のポートIDへ書き換える（ステップ1112）。

【0118】

また、転送先が当該PA内のポート141に接続されたホスト100又は外部ストレージ180の場合は、PAは、当該ポート141のアドレスを当該フレームの送信元アドレスへ設定する（ステップ1113）。

そして、コマンド、データ、受信準備完了報告、完了報告フレームを受信したPAは、決定された転送先であるホスト100、又は外部ストレージ180、CA、AA又は他のPAへ当該フレームを転送する（ステップ1114）。さらに、転送を完了した当該フレームが完了報告フレームだった場合は、PAは、当該入出力処理のルーティング制御情報の登録を解除する（ステップ1115、1116）。

【0119】

また、ステップ1102で、PAが受信したフレームがNASのコマンド、即ちファイルアクセスのコマンドであった場合、PAは、フレームヘッダにセットさ

れているIPアドレスおよびコマンドにセットされているファイル名から得られるディレクトリ情報を元に、NAS管理情報の複製221を検索し、コマンドを処理すべきAAを決定し、当該AAにコマンドを転送する。つぎに、AAはファイルアクセス処理229（ステップ1118）を行う。尚、ファイルアクセス処理229の詳細については、後述する。

【0120】

また、ステップ1102で、PAが受信したフレームがFCPもしくはiSCSI関連のフレームでなく、かつ、NASのコマンドでもなかった場合は、フレームがFCのフレームであればファイバチャネルのノードポートとして従来公知な処理を当該フレーム受信PAの制御プロセッサ142にて実行し、フレームがIPのフレームであればPAはフレームの廃棄処理を行う（ステップ1117）。

【0121】

なお、要求振り分け処理が実行された結果、コマンドの転送先がCA又はAAであると決定された場合、PAは、フレーム以外にアクセス対象であるCA番号またはAA番号、下位論理デバイス番号または上位論理デバイス番号または外部デバイス番号、LBA、サイズなどの情報を付加して転送先のCAもしくはAAへ送信する。また、コマンドに続くデータなどをCAもしくはAAに転送する場合、PA、CA又はAAのそれぞれの転送制御部において、既に確立したコマンドシーケンスを意識したデータ転送を行う。

【0122】

図21は、ファイルアクセス処理229の処理フローの一例を示す図である。要求振り分け処理251の結果、AAはPAからコマンドを受信する（ステップ2101）。AAは、コマンドで要求された処理対象のファイル関連の情報が自身のメモリ149にキャッシュされているかをキャッシュ管理情報230でチェックする（ステップ2102）。処理対象のファイル関連の情報（ディレクトリ情報等）が無かった場合、AAはNAS管理情報の複製227を参照して関連情報が記憶されているデバイス（CAが提供する論理デバイス又は外部デバイス）からディレクトリ情報を読み出しメモリ149へ格納する（ステップ2103）。

【0123】

情報があつた場合、または、情報を読み出しキャッシュに格納した後、AAは、要求がリードもしくはライトであるか否かをチェックする（ステップ2104）。リード又はライトでなかった場合、AAは既存技術を使い当該コマンドの処理を行う（ステップ2118）。

【0124】

要求がリード又はライトであつた場合、AAはデータが自己のメモリ149のキャッシュ領域にあるか否かチェックする（ステップ2105）。要求がリードで、かつ既にデータがメモリ149のキャッシュ領域にあつた場合、AAはそのデータを元にリード処理を行う。

要求がリードで、かつデータがメモリ149のキャッシュ領域に無かつた場合、AAはメモリ149にキャッシュ領域を確保する（ステップ2106）。キャッシュ領域を確保後、もしくは、要求がライトであつた場合、AA自身のアクセスモードが「連携」か否かをチェックする（ステップ2107）。

【0125】

「連携」でなかった場合、AAはコマンドの内容とディレクトリ情報に従い、当該データが記録されているCA又は外部ストレージ180が接続されているポート141があるPAへ入出力要求を送信する（ステップ2108）。送信したコマンドがリード要求であつた場合（ステップ2109）、AAは、コマンド処理254を実施したCAもしくは要求ふり分け処理251を実施したPAからデータを受信し、受信したデータをキャッシュ領域に格納し、その後完了報告を受信（ステップ2110、2113）する。

【0126】

ステップ2109で送信したコマンドがライト要求であつた場合、AAは、コマンド処理254を実施したCAもしくは要求ふり分け処理251を実施したPAから準備完了報告を受信し（ステップ2111）、ついでデータを送信し（ステップ2112）、完了報告を受信する（ステップ2113）。

その後、AAは受信したデータに基づいて、コマンドで要求された操作ができる分のデータのリードを完了したか、またはデータのライトを完了したかをチェックし、完了していなかつた場合、AAはステップ2101～2114の処理を

データのリードまたはライトが完了するまで繰り返す（ステップ2114）。

ステップ2107で連携モードであった場合、AA、PA、SAは、連携モード処理を行う（ステップ2117）。

【0127】

ステップ2105で処理を完了した場合、ステップ2114、ステップ2117及びステップ2118の処理の終了後、AAは、その結果をコマンド応答にしてコマンド送信元のPAへ送信する（ステップ2115）。コマンド応答を受信したPAは、コマンド送信元のホスト100へ当該コマンド応答を送信する（ステップ2116）。

【0128】

なお、本実施形態では、PAでは、TCP/IPまでの処理を行い、NFS等NASのアクセスに使用されるプロトコルにしたがったコマンドがPAとAA間で送受される。

また、本実施形態では、ライト要求に伴うデータを直接CAに書き込む形をとったが、一旦AAのメモリ149のキャッシュ領域に書き込んでから、適当な時にCAに書き込む方式でも良い。また、本実施形態では、AAが、1回以上のリード／ライト要求をCA又は外部ストレージ180に接続されるPAに対して行う方式をとったが、CA、PA及びAAとの間で指示方法を予め決めておき、一回のリード／ライト要求で処理が終了するような方式としても良い。

【0129】

図22は、ファイルアクセス処理229の処理内で実施される連携モード処理2117の処理フローの一例を示す図である。連携処理の場合、AAは、コマンド送信元のPAへ連携モードで動作することを通知する（ステップ2201）。次にAAは、リードまたはライトを行うデバイスの処理を行うCAまたは外部ストレージ180と接続されるポート141があるPAへ連携モードでのリード／ライト要求を送信する（ステップ2202）。

【0130】

送信したのがリード要求であった場合（ステップ2203）、コマンド送信元のPAは、コマンド処理254を実施したCA又は要求ふり分け処理251を実

施したPAからデータを受信（ステップ2204）し、AAは完了報告をコマンド処理254を実施したCA又は要求振り分け処理251を実施したPAから受信（ステップ2205）する。ステップ2203でライト要求であった場合、コマンド送信元のPAは、コマンド処理254を実施したCA又は要求振り分け処理251を実施したPAから準備完了を受信し（ステップ2207）、ついでデータを送信する（ステップ2208）。その後、AAは、その完了報告をコマンド処理254を実施したCA又は要求振り分け処理251を実施したPAから受信する（ステップ2205）。

【0131】

AAは受信したデータでコマンドで要求された操作ができる分のデータのリードを完了したか、またはデータのライトを完了したかをチェックし（ステップ2206）、完了していなかった場合は、ステップ2202～2206の処理をデータのリードまたはライトが完了するまで繰り返す。

【0132】

データリードまたはライトが完了した後、AAは図21の処理フローに戻り、結果をコマンド応答にしてコマンド送信元のPAへ送信し（ステップ2115）、PAは、先にAAから受信した連携モード通知、PAまたはCAから受信したデータ、AAから受信したコマンド応答からコマンド応答を作成し、コマンド送信元のホスト100へ当該コマンド応答を送信する（ステップ2116）。

【0133】

図12は、コマンド処理254の処理フローの一例を示す図である。コマンド処理は、リードやライトの要求を受信したCAにおいて下位論理デバイスに対する入出力要求を処理するものである。

【0134】

要求振り分け処理251の結果又はファイルアクセス処理229の結果、PA又はAAから受信したコマンドについて、CAは、コマンド種別をチェックする（ステップ1202）。当該コマンドの要求がリードの場合、CAは、ディスクキャッシュ154上にリード対象のデータがヒットしているか（キャッシュに格納されているか）をキャッシュ管理情報203を参照して判定する（ステップ1

203)。

【0135】

キャッシュミスの場合、CAはキャッシュ管理情報203を更新してディスクキャッシュ154に領域を確保し(ステップ1204)、当該データが格納される物理デバイス又は外部デバイスからキャッシュへデータを読み出す。リードデータが格納されているデバイスが物理デバイスであるか外部デバイスであるかは、CAが下位論理デバイス管理情報210を参照することによって判断することができる(ステップ1205)。

【0136】

リードデータが格納されているデバイスが物理デバイスであれば、CAは物理デバイス管理情報202より特定したディスク装置157に対してリード要求を発行してデータを読み出し(ステップ1206)、読み出したデータをディスクキャッシュ154の確保領域に格納する(ステップ1207)。

【0137】

リードデータが格納されているデバイスが外部デバイスであれば、CAは外部デバイス管理情報205を参照してエントリ78からアクセス対象PAを特定し、外部デバイスに対するリード要求を当該PAへ送信する(ステップ1210)。

リード要求を受信したPAでは要求振り分け処理251によって当該リード要求を外部ストレージ180に対して送信し、応答として外部ストレージ180からリードデータを受信し、外部ストレージ180から受信したリードデータをCAへ返送する(ステップ1211)。CAは、PAから受信したリードデータをディスクキャッシュ154の確保領域に格納する(ステップ1207)。

【0138】

ディスクキャッシュにデータが存在していた又はディスクキャッシュにリードデータが格納されたら、CAは、ディスクキャッシュ154に格納されたデータをコマンド転送元のPA又はAAへ送信し(ステップ1208)、PAは要求振り分け処理251を実行して当該データをホスト110へ転送し、AAはファイルアクセス処理の続きを行う。

【0139】

CAがPA又はAAから受信したコマンドがライトの場合、CAはライトデータに対応する旧データがディスクキャッシュ154に格納されているか否かを判断し(1212)、キャッシュミスの場合にはディスクキャッシュ154領域を割り当てる(1213)。

【0140】

ステップ1212又は1213の終了後、CAは、コマンド転送元のPA又はAAに対して、転送準備完了を送信する(ステップ1214)。PAへ転送準備完了を送信した場合、転送準備完了を受信したPAは要求振り分け処理251を実行して転送準備完了をホスト100へ転送し、その結果ホスト100からライトデータが送られてくると、PAは転送準備完了を送信したCAへデータを送信する。一方、AAへ転送準備完了を送信した場合、転送準備完了を受信したAAは、CAへデータを送信する。CAではPA又はAAから受信したライトデータを、確保したディスクキャッシュ154上の領域へ格納する(ステップ1215、1216)。

【0141】

また、CAで受信したコマンドがリードでもライトでもない場合、例えばモードセンスなどのセンス系コマンドや診断系のコマンドの場合は、CAは、一般的なSCSIストレージの仕様にに基づき適切な処理を実行する(ステップ1217)。

【0142】

ステップ1208、1217又は1216の処理を終了したCAは、コマンド処理を完了したことをコマンド転送元のホスト100に接続されているPA又はAAへ報告する。完了報告を受領したPAは、要求振り分け処理251を実行して、ホスト100へ完了報告フレームを送信する。完了報告を受領したAAは、ファイルアクセス処理の続きを行う(ステップ1209)。

【0143】

なお、ステップ1201でCAがAAからコマンドを受信した場合でAAが「連携」モードで動作している場合、CAは、AAからはコマンドの他に連携モー

ドの通知と連携するべきPAを指定する情報等を受信する。この場合、CAは、ステップ1208およびステップ1214、1215における準備完了とデータの送受信を、連携を指示されたPAに対して行う。「連携」モードで動作しているPAは、AAからの指示とCAからの準備完了およびデータの送受信に従い、連携モード時のファイルアクセス処理を完了する。

【0144】

図13は、ライトアфта処理257の処理フローの一例を示す図である。ライトアфта処理とは、CAにおけるコマンド処理254の結果、ディスクキャッシュ154に格納されたライトデータをCAがディスク装置157へ書き出す処理である。ディスクキャッシュ154上に保持されたライトデータはキャッシュ管理情報203で管理される。通常、ライトデータやディスクからリードされたリードデータは、なるべく古いものから順にディスクキャッシュ154より追い出されるように、キューなどで管理されている。

【0145】

CAは、キャッシュ管理情報203で管理されたデータの中から、ディスク装置157へ実際に書き込むデータを選択し（ステップ1301）、当該データが格納されるディスク装置157が外部デバイスに対応づけられているか物理デバイスに対応付けられているかを下位論理デバイス管理情報201に基づいて判断する（ステップ1302）。

【0146】

ライトデータを物理デバイスに書き込む場合、CAは物理デバイス管理情報202を参照してデータを書き込むディスク装置157を特定し、ライトデータを特定されたディスク装置157へ転送する（ステップ1303）。

【0147】

一方ライトデータを外部デバイスに書き込む場合には、CAは外部デバイス管理情報205を参照してライトデータが書き込まれる外部デバイスを有する外部ストレージ180と接続されているPAを特定し、当該PAへライト要求を送信する（ステップ1304）。ライト要求を受信したPAは、要求振り分け処理251を実行して当該ライト要求を外部ストレージ180に転送し、ライト要求に

対する応答として当該外部ストレージ180から転送準備完了通知を受信する。すると当該PAは転送準備完了通知をライト要求送信元のCAへ送信する。完了通知を受信したCAはライトデータを当該PAへ送信する（ステップ1305、1306）。

【0148】

ライトデータを受信したPAは、要求振り分け処理251を実行してライトデータを上述の外部ストレージ180に送信し、外部ストレージ180によってライトデータが外部デバイスに書き込まれる。

【0149】

ライトデータを物理デバイスもしくは外部デバイスへライトした後、CAは、ライトデータが格納されていたディスクキャッシュ154上の領域を解放する（ステップ1307）。

【0150】

次に、AA又はPAのポート141に障害が発生した場合のストレージ130の動作について説明する。AA又はPAのポート141に障害が発生した場合、ストレージ130は、代替のAA又はポート141を用いて処理を継続する。

【0151】

図20は、NAS／ポート障害処理231の処理フローの一例を示す図である。ポート141もしくはAAの障害を検出または他の部位から通知されたMAは、検出もしくは通知された障害がポート障害かAA障害かを判断する（ステップ2001、2002）。検出等された障害がポート障害であった場合、MAは、障害が発生したポート141のポート管理情報224を参照し、当該ポート141の管理情報のエントリ1705を「障害」に設定する。

【0152】

次に、MAは、エントリ1708に登録されている代替ポートのポート管理情報に障害の発生したポート141の情報を追加コピーする。代替ポートのエントリ1705が「通信」状態の場合は、障害の発生したポート141のエントリ1704に登録されていたIPアドレスリストを代替ポートのIPアドレスリストに追加する。「通信」状態でなかった場合、MAは、障害の発生したポート141の

MACアドレス、IPアドレスリスト、アクセスモード等を代替ポートに対応するポート管理情報の各エントリにコピーする。そして、MAはPAにポート管理情報の更新を指示する（ステップ2003）。

【0153】

AA障害であった場合、MAは障害が発生したAAに対応するNAS管理情報223を参照し、当該管理情報223のエントリ1603を「障害」に設定する。次に、MAは障害が発生したAAに関するNAS管理情報223のエントリ1606に登録されている代替AAのNAS管理情報223に、障害の発生したAAのIPアドレス、ディレクトリ名、対応CA番号、上位論理デバイス／外部デバイスのリストを追加する。その後、MAは、AAとPAにNAS管理情報223の更新を指示する（2004）。

【0154】

MA、PA及びAAでポート管理情報とNAS管理情報の更新が終了した後、MAは、STにポート管理情報とNAS管理情報の更新を通知する（ステップ2005）。ポート管理情報とNAS管理情報の更新を通知されたSTは、更新のあった情報を取り込んだ後、要求元であるストレージ管理者もしくは管理サーバへ報告すべく完了報告を出力する（ステップ2006）。

【0155】

図18は、NAS管理情報変更処理225の処理フローの一例を示す図である。なお、新規にNAS管理情報を登録する場合、および、あるAAにかかわるNAS管理情報の一部もしくは全てを別AAにコピーして処理を引き継がせる場合も本処理が実行される。更に本処理は、AAおよびPAが各自の処理負荷を測定する機構を持ち、その処理負荷の情報をMAもしくはST経由で管理サーバが収集し、その情報を元に、ストレージ管理者または、管理サーバ上の記載しない処理負荷管理処理部が処理負荷の高いAAもしくはAAの処理にかかわるPAから、処理負荷の低いAAもしくはPAへ負荷を分散させるように制御する時に動作する。

【0156】

また、一部のAAの修理や、AAを高性能なものに変更する等のために新しい物と交換するような場合、ストレージ管理者または、管理サーバからの制御で動

作させる。

【0 1 5 7】

ストレージ管理者もしくは管理サーバ 1 1 0 から NAS 管理情報変更指示を受け付けた S T は、M A に対して同指示を転送する（ステップ 1 8 0 1）。同指示には、情報を変更する A A にかかわる NAS 管理情報エントリが付加される。指示を受信した M A は、NAS 管理情報を変更すると同時に、変更する NAS 管理情報のエントリにかかわる A A に対し、NAS 管理情報を変更するように指示を送信する（ステップ 1 8 0 2）。指示を受信した A A は、NAS 管理情報の複製 2 2 7 の変更を行う（ステップ 1 8 0 3）。変更指示が新たなエントリの追加であった場合（ステップ 1 8 0 4）、M A に登録の完了を通知する（ステップ 1 8 0 7）。

【0 1 5 8】

変更指示が既存エントリの変更であった場合、A A は、当該変更にかかわる処理中のコマンドを廃棄し（ステップ 1 8 0 5）さらに、メモリ 1 4 9 にキャッシュされている当該変更にかかわるディレクトリ情報をクリアして（ステップ 1 8 0 6）、M A に登録の完了を通知する（ステップ 1 8 0 7）。A A からの登録完了通知を受信した M A は、NAS 管理情報の更新があったことを各 P A 及び S T に通知する（ステップ 1 8 0 8）。

【0 1 5 9】

NAS 管理情報更新を通知された P A では、更新のあった NAS 管理情報の複製 2 2 1 をメモリ 1 4 3 へ取り込む。又、NAS 管理情報更新を通知された S T では、更新のあった情報を取り込んだ後、要求元であるストレージ管理者もしくは管理サーバ 1 1 0 へ報告すべく完了報告を出力する（ステップ 1 8 0 9）。

【0 1 6 0】

尚、N A S / ポート障害処理にて NAS 管理情報が変更される際には、上述した処理手順のうち、ステップ 1 8 0 2 から処理が開始される。

【0 1 6 1】

図 1 9 は、ポート管理情報変更処理 2 2 6 の処理フローの一例を示す図である。なお、新規にポート管理情報を登録する場合及びあるポート 1 4 1 にかかわるポート管理情報の一部もしくは全てを別ポート 1 4 1 にコピーして処理を引き継

がせる場合も、本処理が実行される。更に本処理は、P Aが各自の処理負荷を測定する機構を持ち、その処理負荷の情報をMAもしくはS T経由で管理サーバが収集し、その情報を元に、ストレージ管理者または、管理サーバ上の記載しない処理負荷管理処理部が処理負荷の高いポートから、処理負荷の低いポートへ負荷を分散させるように制御する時に動作する。また、一部のP Aの修理や、P Aを高性能なものに変更する等のために新しい物と交換するような場合、ストレージ管理者または、管理サーバからの制御で動作させる。

【0162】

ストレージ管理者もしくは管理サーバ110からNAS管理情報変更指示を受け付けたS TがMAに対して同指示を転送する。同指示には、情報を変更するポートにかかわるポート管理情報のエントリが付加される（ステップ1901）。指示を受信したMAは、ポート管理情報を変更すると同時に、変更するポート管理情報のエントリにかかわるポート141が存在するP Aに対し、ポート管理情報の複製222を変更するように指示を送信する（ステップ1902）。

【0163】

指示を受信したP Aは、ポート管理情報の複製222の変更を行う（ステップ1903）。変更指示が新たなエントリの追加または、既存エントリへの情報の付加であった場合（ステップ1904）、P AはMAに登録の完了を通知し、更新後のポート管理情報の複製222の内容に従い自P A内のポート141で通信処理を開始する（ステップ1907）。変更指示が既存エントリの内容の変更であった場合、P Aは当該変更にかかわる処理をクリアし、更新後のポート管理情報の複製222の内容に従い自P A内のポート141で通信処理を開始した（ステップ1905）後、MAに登録の完了を通知する（ステップ1907）。P Aからの登録完了通知を受信したMAは、ポート管理情報の更新があったことをS Tに通知する（ステップ1908）。

【0164】

ポート管理情報更新を通知されたS Tは、更新のあった情報を取り込んだ後、要求元であるストレージ管理者もしくは管理サーバ110へ報告すべく完了報告を出力する（ステップ1909）。

【 0 1 6 5 】

尚、N A S / ポート障害処理にてポート管理情報が変更される際には、上述した処理手順のうち、ステップ 1 9 0 2 から処理が開始される。

【 0 1 6 6 】

次に、本システムの各部位で動作するストレージ制御処理について説明する。ストレージ 1 3 0 内に存在する物理デバイス、及び外部ストレージ 1 8 0 に存在する外部デバイスを含むデバイスを、特定のホスト 1 0 0 または A A に割り当てて当該ホスト 1 0 0 または A A から使用できるようにする処理は、外部デバイス定義処理 2 5 3、論理デバイス定義処理 2 5 5、L U パス定義処理 2 5 2 の大きく 3 つに分けることができる。システム管理者は、上述の 3 つの処理を順に行うことで、上位論理デバイスをホスト 1 0 0 又は A A に割り当てるようにストレージ 1 3 0 を設定することができる。以下、この 3 つの処理について説明する。

【 0 1 6 7 】

図 8 は、外部デバイス定義処理 2 5 3 の処理フローの一例を示す図である。外部デバイス定義処理 2 5 3 は、ストレージ 1 3 0 の管理下に外部ストレージ 1 8 0 内のデバイスを外部デバイスとして導入するための処理である。

【 0 1 6 8 】

まず、ストレージ管理者もしくは管理サーバ 1 1 0 からの外部ストレージ 1 8 0 接続指示を受け付けた S T が、M A に接続指示を送信する。接続指示には、接続対象となる外部ストレージ 1 8 0 を特定する情報、例えば、外部ストレージ 1 8 0 のポート 1 8 1 の WWN、IP アドレス、又は外部ストレージデバイスへ I n q u i r y コマンドを送信した際にその応答により得られる装置識別情報もしくはその両方の情報及び外部ストレージ 1 8 0 に接続されるストレージ 1 3 0 のポート 1 4 1 の番号とが付加される（ステップ 8 0 1）。

【 0 1 6 9 】

外部ストレージの接続指示を受信した M A は、接続指示に付加された全てのポート 1 4 1 の番号に対応する全 P A へ、外部ストレージ 1 8 0 を識別するための情報と外部ストレージ 1 8 0 と接続されるポート 1 4 1 の番号とを有する外部ストレージ接続指示を送信する（ステップ 8 0 2）。

【0170】

MAから外部ストレージ接続指示を受信したPAは、接続指示に付加された外部ストレージ180の識別情報（以下「外部ストレージ識別情報」）を用いて、接続対象となる外部デバイスを探索する。具体的には、外部ストレージ識別情報としてポート181のWWNまたはIPアドレスを受信した場合、PAは接続指示に付加されたポート番号により指定されたポート141から外部ストレージ180のポート181に対応する全LUNに対してInquiryコマンドを送信し、Inquiryコマンドに対する正常な応答を返したLUNを外部デバイス登録候補とする。

【0171】

外部ストレージ識別情報として装置識別情報しか受信しない場合、PAは、当該PAの全ポート141各々によって検出された外部ストレージ180のノードポート（ノードポートログイン時に検出済み）のそれぞれに対応する全LUNにInquiryコマンドを送信する。その後、Inquiryコマンドに対する正常な応答を返したデバイス（LUN）について、PAは応答内に含まれる装置識別情報を外部ストレージ識別情報と比較し、一致するデバイスを外部デバイス登録候補とする（ステップ803）。

【0172】

PAは、検出された外部デバイス登録候補についての情報リストをMAに返却する。この情報リストには、各外部デバイス登録候補について外部デバイス管理情報205を設定するのに必要な情報が含まれる（ステップ804）。

【0173】

情報リストを受取ったMAは、受け取った外部デバイスリストに含まれる外部デバイスに採番し、外部デバイス管理情報205へデバイス情報を登録し、同情報に更新のあった旨を他の各部位へ通知する。なお、本処理ステップにおいてMAが外部デバイス管理情報205へ登録する情報には、具体的には、MAが外部デバイスに対して割り当てた外部デバイス番号、Inquiryコマンドに対する応答から取得されるサイズ、ストレージ識別情報及び外部ストレージ内デバイス番号、PAからMAに通知されるPA/イニシエータポート番号リスト及びタ

ーゲットポート ID/ターゲット ID/LUN リストが含まれる。

【0174】

また、対応上位論理デバイス番号及び対応 CA 番号/下位論理デバイス番号/A 番号は未割当のため、MA は初期値である無効値を対応するエントリに設定する。更にエントリ 75 には「オフライン」状態が設定される（ステップ 805）。

【0175】

外部デバイス管理情報 205 の更新通知を受けた各 CA、PA 及び AA は、MA の制御メモリ 164 に保存されている外部デバイス管理情報 205 を各々のメモリに読み込む。また、ST は外部デバイス管理情報 205 をメモリに取り込むと共に、外部デバイス定義処理の完了を、当該処理の要求元であるストレージ管理者もしくは管理サーバ 110 に通知すべく完了通知を出力する（ステップ 806）。

【0176】

なお、ストレージ 130 に対してストレージ管理者もしくは管理サーバ 110 が接続指示と共に、導入対象となる外部ストレージ 180 を指示せず、外部ストレージ 180 の接続指示のみをストレージ 130 に指示し、ストレージ 130 で全ポート 141 から検出した全外部ストレージ 180 各々が有する全デバイスを外部デバイスとして登録しても構わない。また、ストレージ管理者や管理サーバ 110 からは特に明示的な接続指示を与えず、ストレージ 130 に外部ストレージ 180 が接続されたのを契機に、ストレージ 130 が検出した全デバイスを外部デバイスとして登録してもよい。

【0177】

図 9 は、論理デバイス定義処理 255 の処理フローの一例を示す図である。論理デバイス定義処理 255 は、ストレージ管理者もしくは管理サーバ 110 からの指示を受けて、ストレージ 130 で定義される物理デバイス又は外部デバイス定義処理 253 で定義された外部デバイスに対して、上位又は下位論理デバイスを定義する処理である。

【0178】

ストレージ管理者もしくは管理サーバ 1 1 0 から論理デバイス定義指示を受け付けた S T は、当該指示を M A に送信する。論理デバイス定義指示には、論理デバイスが割り当てられる対象となる物理デバイス又は外部デバイスの番号、割り当てられる上位論理デバイスの番号、割り当てられる下位論理デバイスの番号及び当該下位論理デバイスを管理する C A の番号とが付加される。なお本実施形態では、説明の簡略化のため、1 つの物理もしくは外部デバイスに対して 1 つの論理デバイスを割り当てるものとするが、2 つ以上の物理もしくは外部デバイスからなるデバイスグループに対して 1 つの論理デバイスを、又は 1 つの物理もしくは外部デバイスに対して 2 つ以上の論理デバイスを、又は 2 つ以上の物理もしくは外部デバイスからなるデバイスグループに対して 2 つ以上の論理デバイスを定義しても構わない。

【0 1 7 9】

ただし、それぞれの場合、下位論理デバイス管理情報 2 0 1 に、物理もしくは外部デバイス内での当該論理デバイスの開始位置およびサイズなどの情報の付加、対応する物理および外部デバイス番号エントリの複数デバイス化、物理および外部デバイス管理情報 2 0 2、2 0 5 に、対応する下位論理デバイス番号エントリの複数デバイス化を行うことがそれぞれ必要となる（ステップ 9 0 1）。

【0 1 8 0】

論理デバイス定義指示を受信した M A は、論理デバイス定義指示に付加されている情報より対象 C A を特定し、特定された C A に論理デバイス定義指示を送信する（ステップ 9 0 2）。

【0 1 8 1】

論理デバイス定義指示を受信した C A は、指示された物理もしくは外部デバイスに対して下位論理デバイスを登録する（ステップ 9 0 3）。具体的に C A は、下位論理デバイス管理情報 2 0 1 の対象デバイスエントリ（論理デバイス定義指示に付加されている下位論理デバイスの番号に対応するエントリ）について、論理デバイス定義指示に付加されている下位論理デバイス番号をエントリ 5 1 に、論理デバイス定義指示に付加されている物理もしくは外部デバイスの番号をエントリ 5 3 に、下位論理デバイスのサイズをエントリ 5 2 に、論理デバイス定義指

示に付加されている上位論理デバイス番号をエントリ 55 に各々登録し、エントリ 54 には「オンライン」を設定する。

【0182】

また、当該物理／外部デバイスの対応 CA 番号／下位論理デバイス番号を設定し、デバイス状態を「オンライン」に更新する。登録が完了したら CA は MA にその旨通知する（ステップ 903）。

【0183】

次に MA は、論理デバイス定義指示によって指定された下位論理デバイスに上位論理デバイスを割り当て、制御情報が更新された旨を他の各部位に通知する。具体的に MA は、上位論理デバイス管理情報 203 の当該デバイスエントリ（論理デバイス定義指示に付加されている上位論理デバイスの番号に対応するエントリ）について、論理デバイス定義指示に付加されている上位論理デバイスの番号をエントリ 31 に、上位論理デバイスのサイズをエントリ 32 に、論理デバイス定義指示に付加されている CA 番号と下位論理デバイス番号をエントリ 33 に設定し、デバイス状態としてエントリ 34 には「オフライン」状態を設定し、エントリ 35 及び 36 に関しては未割当のため無効値を設定する。

【0184】

更に、当該上位論理デバイスの割当対象が外部デバイスの場合には、論理デバイス定義指示に付加されている外部デバイス番号をエントリ 38 に設定し、デバイスアクセスモードを初期値として「PA 直結」に設定する。更に、当該上位論理デバイスの割り当て対象が外部デバイスである場合には、外部デバイス管理情報 205 の対応エントリ 73 に当該上位論理デバイス番号を設定する。そして、MA は制御情報の更新があった旨を各 PA 及び ST に通知する（ステップ 904）。

【0185】

制御情報更新を通知された PA は更新のあった情報をメモリ 143 へ取り込む。又、制御情報更新を通知された ST は、更新のあった情報を取り込んだ後、論理デバイス定義処理の完了を要求元であるストレージ管理者若しくは管理サーバ 110 へ報告すべく終了報告を出力する（ステップ 905）。

【0186】

図10は、LUパス定義処理252の処理フローの一例を示す図である。ストレージ管理者もしくは管理サーバ110からLUパス定義指示を受け付けたSTがMAに対して同指示を転送する。同指示には、対象上位論理デバイス番号、LUを定義するポート141の番号、LUNに加えて、同LUにアクセスするホスト100の識別情報（ホストが有するポート107のWWNなど）またはAAの識別情報（AA番号等）及び当該上位論理デバイスへの要求機能レベルを示す情報が付加される。

【0187】

ここで要求機能レベルとは、上位論理デバイスに要求されるデータの信頼性等のレベルを指す。具体的には、ストレージ130が提供するデータ複製やデバイス再配置などのデータ連携機能や、データをキャッシュ上に格納しておくキャッシュ常駐化によるアクセス高速化機能などの適用有無が指定される。これによって、機能の適用の有無により、データの信頼性のレベルが変化する。なお、要求機能レベルとして、単に、当該上位論理デバイスに格納されるデータがCA経由で処理が必要か否か（たとえ、CAに接続される記憶装置157にデータが格納されなくても）を示す情報を指定してもよい（ステップ1001）。

【0188】

LUパス定義指示を受信したMAは、LUパス定義指示で指定される上位論理デバイスに関する上位論理デバイス管理情報203内エントリ38より、当該上位論理デバイスに対応するデバイスが外部デバイスか否かを判定し（ステップ1002）、外部デバイスの場合は外部デバイス接続変更処理を行う（ステップ1003）。

【0189】

外部デバイス接続変更処理完了後又はLUパス定義指示で指定される上位論理デバイスが外部デバイスに対応付けられていない場合、MAは、当該上位論理デバイスに対してLUパス登録を行う。具体的にMAは、上位論理デバイス管理情報203の当該デバイスエントリのエントリ35及びエントリ36にLUパス定義指示で指定される対応情報を設定し、LUパス管理情報206のLUパス定義

指示で指定されるポート 141 に関する空きエントリに、ターゲット ID/LUN を始めとする構成情報を、LU パス定義指示に従って設定する。登録及び設定が完了したら、MA はその旨を他の各部位へ通知する（ステップ 1004）。

【0190】

通知を受信した PA は、新たに設定、登録された情報を取込み、通知を受信した ST は新たに設定、登録された情報を取り込むと共に、要求元である管理サーバ 110 やストレージ管理者へ完了報告を通知すべく完了報告を出力する（ステップ 1005）。

【0191】

図 14 は、外部デバイス接続変更処理 256 の処理フローの一例を示す図である。外部デバイスに対応する上位論理デバイスに対して LU パス定義される際に、当該外部デバイスに対してストレージ管理者が要求する機能レベルに基づき CA 経由での入出力処理が必要であると判断された場合又は MA が ST を介してストレージ管理者もしくは管理サーバ 110 から上位論理デバイスに対する機能要求レベルの変更指示を受信した場合に、初期状態の PA 直結接続から接続形態を切り替えるため、外部デバイス接続変更処理 256 が実行される。

【0192】

機能要求レベルの変更指示を受信した MA は、当該変更指示内の情報によって特定される上位論理デバイスの接続形態を CA 経由とすべきか否かを、当該変更指示に基づいて判断する（ステップ 1401）。CA 経由とすべきである場合には、MA は、当該上位論理デバイスのアクセスモードが PA 直結モードとなっているかどうかを上位論理デバイス管理情報 203 を参照して判断する（ステップ 1402）。

【0193】

該当する上位論理デバイスが PA 直結モードである場合には、MA は当該上位論理デバイスの接続形態を「PA 直結」から「CA 経由」へ変更する（ステップ 1403）。既に CA 経由モードとなっている場合には変更は必要ないので処理を終了する。

【0194】

ステップ1401で当該上位論理デバイスの接続形態をPA直結とすべき場合、MAは上位論理デバイス管理情報203を参照して当該上位論理デバイスのデバイスアクセスモードがCA経由となっているかPA直結となっているかを判断する（ステップ1404）。

【0195】

当該上位論理デバイスのデバスアクセスモードがCA経由となっている場合、MAは接続形態を「CA経由」から「PA直結」へ変更する必要がある。そこでまずMAは、当該上位論理デバイスの上位論理デバイス管理情報203内のエントリ37を「PA直結移行中」へ変更し（ステップ1405）、当該上位論理デバイスに対応付けられている下位論理デバイスを管理しているCAを上位論理デバイス管理情報203を参照して特定する。そして特定されたCAに当該上位論理デバイスの接続形態がCA経由からPA直結に変更される旨を通知する（ステップ1406）。

【0196】

通知を受けたCAは、対応する下位論理デバイスの全データについてディスクキャッシュ154を検索しディスクキャッシュをミス化、すなわち、物理デバイス又は外部デバイスへ未更新のデータをディスクキャッシュ154から対応デバイスへ書き出してデバイス内のデータを更新した後、更新済みのデータについては即、当該データに割り当てられたキャッシュ領域を解放する。CAは、ディスクキャッシュ154の検索をキャッシュ管理情報を用いて、当該下位論理デバイスの先頭から順次行う。又、CAは、検索およびミス化処理が終わった部分については、上位論理デバイス及び下位論理デバイス管理情報の切替進捗ポイントを進めることで進捗を管理する（ステップ1407）。

【0197】

ディスクキャッシュ154の全領域についてミス化処理が完了したら、CAはその旨をMAへ報告し（ステップ1408）、MAで上位論理デバイス管理情報203内のエントリ37を「PA直結」へ変更して接続形態変更処理を完了する。

【0198】

以上の実施形態によれば、ホスト 1 0 0 がコマンドを送付するアドレスが付与される I/F 部分 (P A) とファイルアクセス処理を行う NAS サーバ部分 (A A) とデバイスの処理を行う キャッシュ制御部 (C A) と外部ストレージ 1 8 0 と接続される I/F 部分 (P A) を分離し、相互結合網 1 7 0 で接続することで、処理負荷を分散可能な NAS システムが実現できる。

【 0 1 9 9 】

また、本実施形態によれば、ホスト 1 0 0 がコマンドを送付するアドレスが付与される I/F 部分 (P A) とファイルアクセス処理を行う NAS サーバ部分 (A A) とデバイスの処理を行う キャッシュ制御部 (C A) と外部ストレージ 1 8 0 と接続される I/F 部分 (P A) とで「連携」モードを設け、大量のデータがホスト 1 0 0 と A A で送受されるような場合、対ホストの I/F 部分 (P A) と キャッシュ制御部 (C A) もしくは外部ストレージ 1 8 0 と接続される I/F 部分 (P A) 間で直接データを送受することができ、A A の処理負荷を軽減できると同時に、ホストに対するコマンド応答が早い NAS システムが実現できる。

【 0 2 0 0 】

また、本実施形態によれば、NAS 管理情報 2 2 3 で、A A が管理するファイルとそれにアクセスするためにホスト 1 0 0 がコマンドを送信する先である I/F 部分の IP アドレスの対応関係を、1 つの P A に対し複数の A A が処理するようにエントリ 1 6 0 2 に設定でき、ホスト 1 0 0 からは 1 台の NAS システムにアクセスするように見えながら、複数の A A で処理を実施することで処理負荷を軽減し、ホスト 1 0 0 に対するコマンド応答が早い NAS システムが実現できる。

【 0 2 0 1 】

また、本実施形態によれば、NAS 管理情報 2 2 3 で A A が管理するファイルとそれにアクセスするためにホスト 1 0 0 がコマンドを送信する先である I/F 部分の IP アドレスの対応関係を、1 つの A A が複数の IP アドレスを持っているようにエントリ 1 6 0 2 に設定でき、ホスト 1 0 0 からは複数の NAS システムが存在するように見えながら、複数の P A またはポート 1 4 1 で I/F 処理を実施することで、処理負荷を軽減し、ホスト 1 0 0 に対するコマンド応答が早い NAS システムが実現できる。

【0202】

また、本実施形態によれば、NAS管理情報223でAAが管理するファイルとそれにアクセスするためにホスト100がコマンドを送信する先であるI/F部分のIPアドレスの対応関係を、1つのAAが複数のIPアドレスを持っているようにエントリ1602に設定でき、ホスト100からは複数のNASシステムが存在するように見え、NASシステム毎に用途を使い分けることができるNASシステムが実現できる。

【0203】

また、本実施形態によれば、NAS管理情報223でAAが管理するファイルとそれにアクセスするためにホスト100がコマンドを送信する先であるI/F部分のIPアドレスの対応関係の一部を他のAAに移すことができ、これにより、ホスト100に影響を及ぼすことなく、処理負荷が大きくなってきたAAを増設し処理負荷を軽減し、ホストに対するコマンド応答を向上できるNASシステムが実現できる。

【0204】

また、本実施例によれば、NAS管理情報223でAAが管理するファイルとそれにアクセスするためにホスト100がコマンドを送信する先であるI/F部分のIPアドレスの対応関係の一部を変更することができ、これにより、処理負荷が大きくなってきたPAを増設もしくは処理負荷が少ないPAへ負荷を分散させ、これにより処理負荷を軽減し、ホストに対するコマンド応答を向上できるNASシステムが実現できる。

【0205】

また本実施形態によれば、NAS管理情報223のAAに関するエントリ全てを新規のAAのエントリにコピーすることで、ホスト100に影響を及ぼすことなく、AAを処理能力の高い新しいパッケージに交換でき、処理負荷を軽減し、ホストに対するコマンド応答を向上できるNASシステムが実現できる。

【0206】

また本実施形態によれば、ポート管理情報224のポートに付与されているIPアドレスの一部を別のポート141のエントリに移動することで、ホスト100

の設定を変更することなく、ポート 141 へのアクセスを分散させ、処理負荷を軽減し、ホストに対するコマンド応答を向上できる NAS システムが実現できる。

【0207】

また本実施形態によれば、AA に障害が発生した場合に、NAS 管理情報 223 に予め設定してあった代替 AA へ当該 AA の管理情報を引き継ぐことができ、ホスト 100 の設定を変更することなく処理を継続可能な NAS システムが実現できる。

【0208】

また本実施形態によれば、PA のポート 141 に障害が発生した場合に、ポート管理情報 224 の当該ポート 141 に関する情報を全てを使用されていないポート 141 のエントリにコピーするか、使用されているポート 141 に当該ポートに関する情報の一部を移動することで、ホスト 100 の設定を変更することなく使用できる NAS システムが実現できる。

【0209】

また、本実施形態によれば、CA が外部ストレージを自デバイスとして AA に提供できるので、ストレージ 130 にディスク装置 157 を搭載せずに、安い外部ストレージを利用することができ、前述の本実施形態の効果を持った低価格の NAS システムを実現することができる。

【0210】

また、本実施形態によれば、CA を介さずに、AA が直接外部ストレージを使用することができ、前述の本実施形態の効果を持った低価格の NAS システムを実現することができる。

【0211】

【発明の効果】

本発明によれば、NAS の使用負荷が上がった場合にも、レスポンスが低下しないように、処理負荷が高い部分の処理を分散するよう制御可能である。

【0212】

また、本発明によれば、障害が発生した場合にも障害発生部分の情報を他に引き継ぐことができ、計算機から当該 NAS へのアクセスを継続もしくは計算機の設

定を変更しないで継続利用できる。

【図面の簡単な説明】

【図 1】

システムのハードウェア構成の一例を示す図である。

【図 2】

ストレージ 1 3 0 のソフトウェア構成の一例を示す図である。

【図 3】

上位論理デバイス管理情報の一例を示す図である。

【図 4】

L U パス管理情報の一例を示す図である。

【図 5】

下位論理デバイス管理情報の一例を示す図である。

【図 6】

物理デバイス管理情報の一例を示す図である。

【図 7】

外部デバイス管理情報の一例を示す図である。

【図 8】

外部デバイス定義処理の一例を示す図である。

【図 9】

論理デバイス定義処理の一例を示す図である。

【図 1 0】

L U パス定義処理の一例を示す図である。

【図 1 1】

要求振り分け処理の一例を示す図である。

【図 1 2】

コマンド処理の一例を示す図である。

【図 1 3】

ライトアフタ処理の一例を示す図である。

【図 1 4】

外部デバイス接続変更処理の一例を示す図である。

【図 1 5】

管理端末のソフトウェア構成の一例を示す図である。

【図 1 6】

NAS管理情報の一例を示す図である。

【図 1 7】

ポート管理情報の一例を示す図である。

【図 1 8】

NAS管理情報変更処理の一例を示す図である。

【図 1 9】

ポート管理情報変更処理の一例を示す図である。

【図 2 0】

NAS／ポート障害処理の一例を示す図である。

【図 2 1】

ファイルアクセス処理の一例を示す図である。

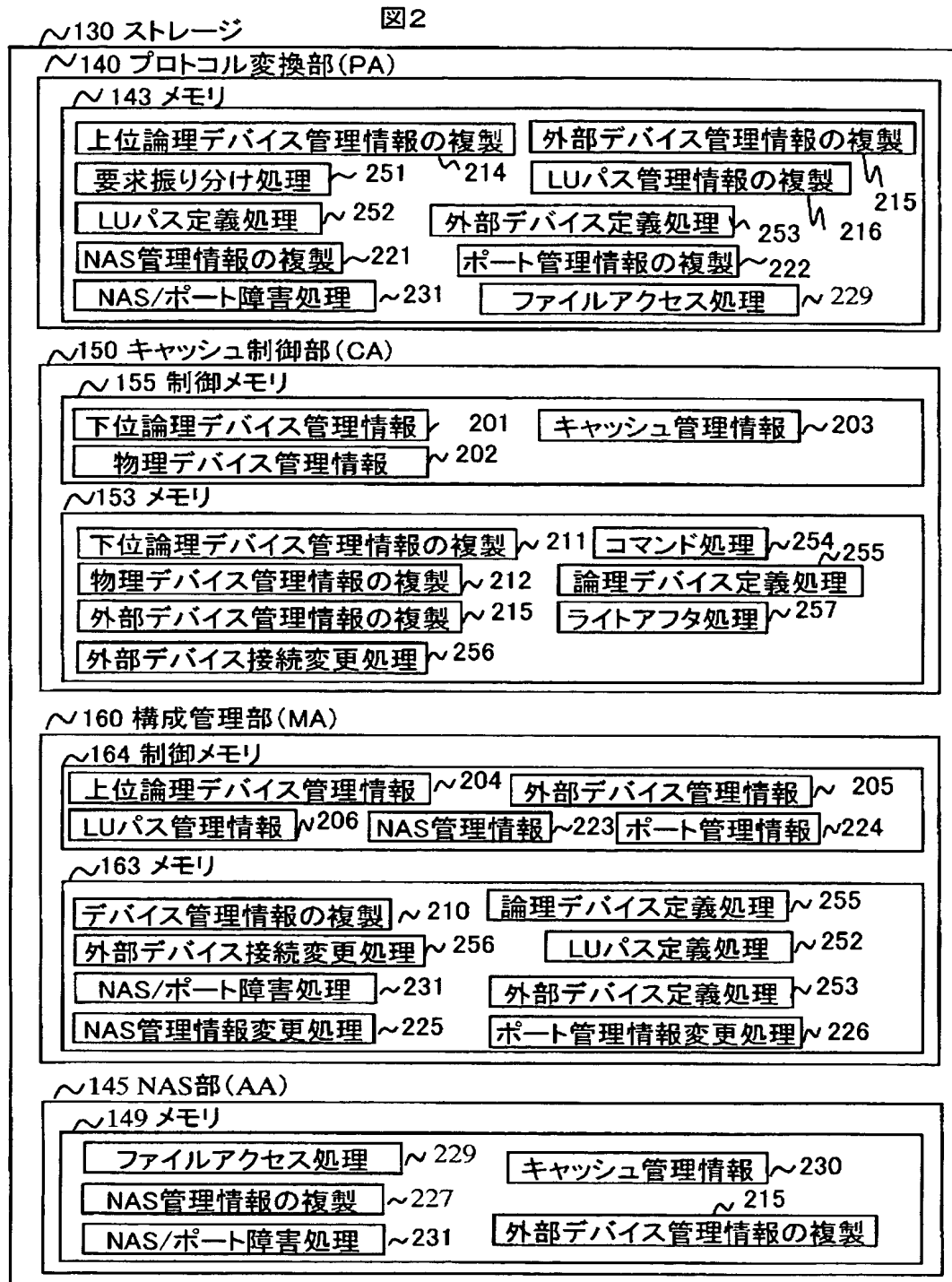
【図 2 2】

連携モード処理の一例を示す図である。

【符号の説明】

1 0 0…ホスト、1 1 0…管理サーバ、1 2 0…スイッチ、1 3 0…ストレージ装置、1 4 0…プロトコル変換部、1 4 5…NAS部、1 5 0…キャッシュ制御部、1 6 0…構成管理部、1 7 0…相互結合網、1 8 0…外部ストレージ装置、1 9 0…管理端末。

【図 2】



【図 3】

図3

上位論理デバイス管理情報204

上位論理デバイス#0の構成情報	上位論理デバイス番号	~31
	サイズ	~32
	対応CA番号、下位論理デバイス番号	~33
	デバイス状態	~34
	ポート番号、ターゲットID、LUN	~35
	接続ホスト名	~36
	デバイスアクセスモード	~37
	対応外部デバイス番号	~38
	切替進捗ポインタ	~39
	⋮	

【図 4】

図4

LUパス管理情報206

ポート/AA番号 #0の構成情報	ターゲットID/LUN	~41
	対応上位論理デバイス番号	~42
	接続ホスト名	~43
	⋮	

【図 5】

図5

下位論理デバイス管理情報201

下位論理デ バイス#0の 構成情報	下位論理デバイス番号	h51
	サイズ	h52
	対応物理/外部デバイス番号	h53
	デバイス状態	h54
	対応上位論理デバイス番号	h55
	切替進捗ポインタ	h56
	⋮	

【図 6】

図6

物理デバイス管理情報202

物理デバイス#0 の構成情報	物理デバイス番号	h61
	サイズ	h62
	対応下位論理デバイス番号	h63
	デバイス状態	h64
	RAID構成(RAIDレベル, データ/パリティ数)	h65
	ストライプサイズ	h66
	ディスク番号リスト	h67
	ディスク内開始オフセット	h68
	ディスク内サイズ	h69
	⋮	

【図 7】

図7

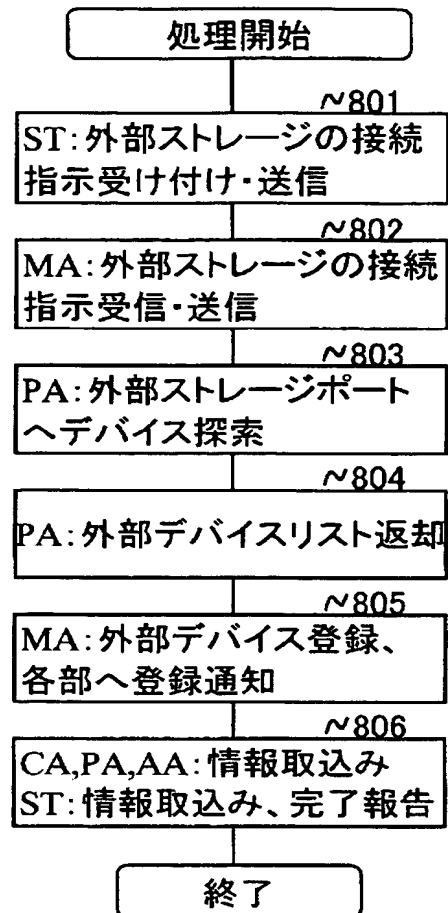
外部デバイス管理情報205

外部デバイス#0
の構成情報

外部デバイス番号	ゝ71
サイズ	ゝ72
対応上位論理デバイス番号	ゝ73
対応CA番号／下位論理デバイス番号／AA番号	ゝ74
デバイス状態	ゝ75
ストレージ識別情報	ゝ76
外部ストレージ内デバイス番号	ゝ77
PA番号／イニシエータポート番号リスト	ゝ78
ターゲットポートID／ターゲットID／LUNリスト	ゝ79
⋮	

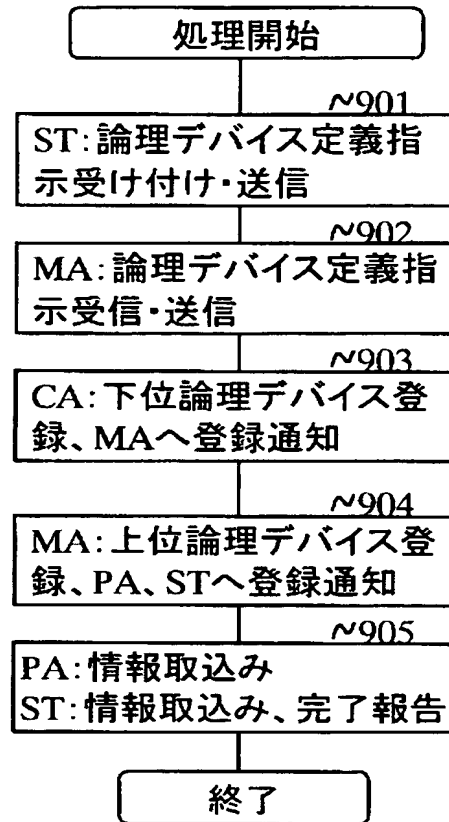
【図 8】

図 8

外部デバイス定義処理253

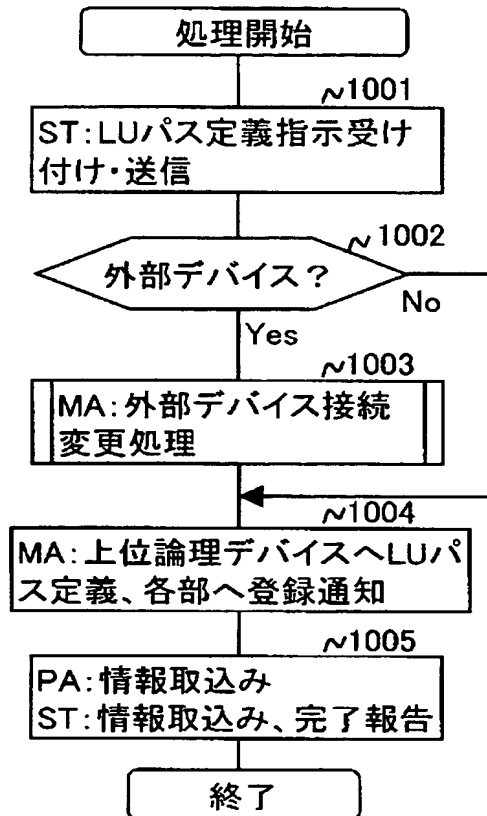
【図 9】

図9

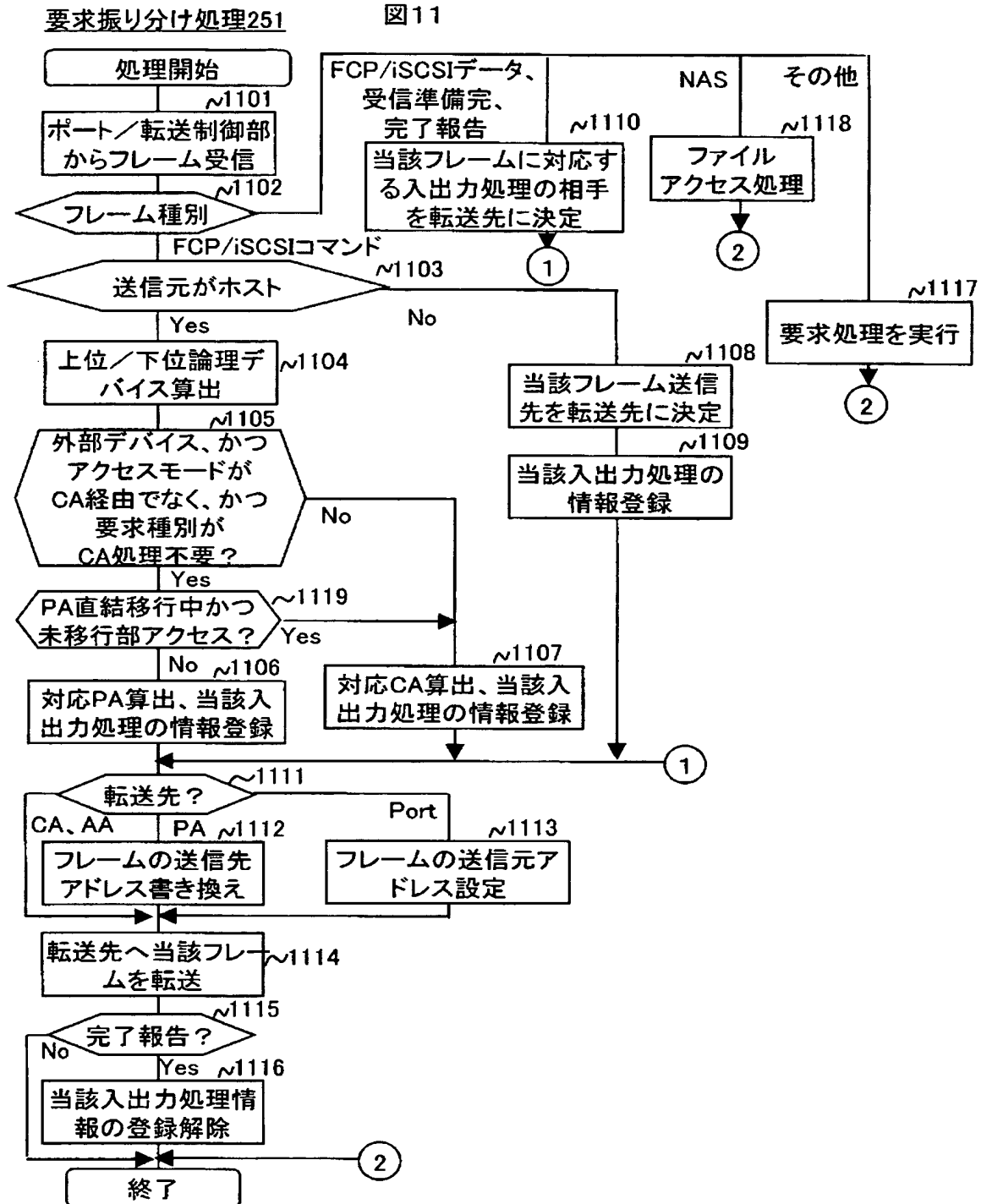
論理デバイス定義処理255

【図 10】

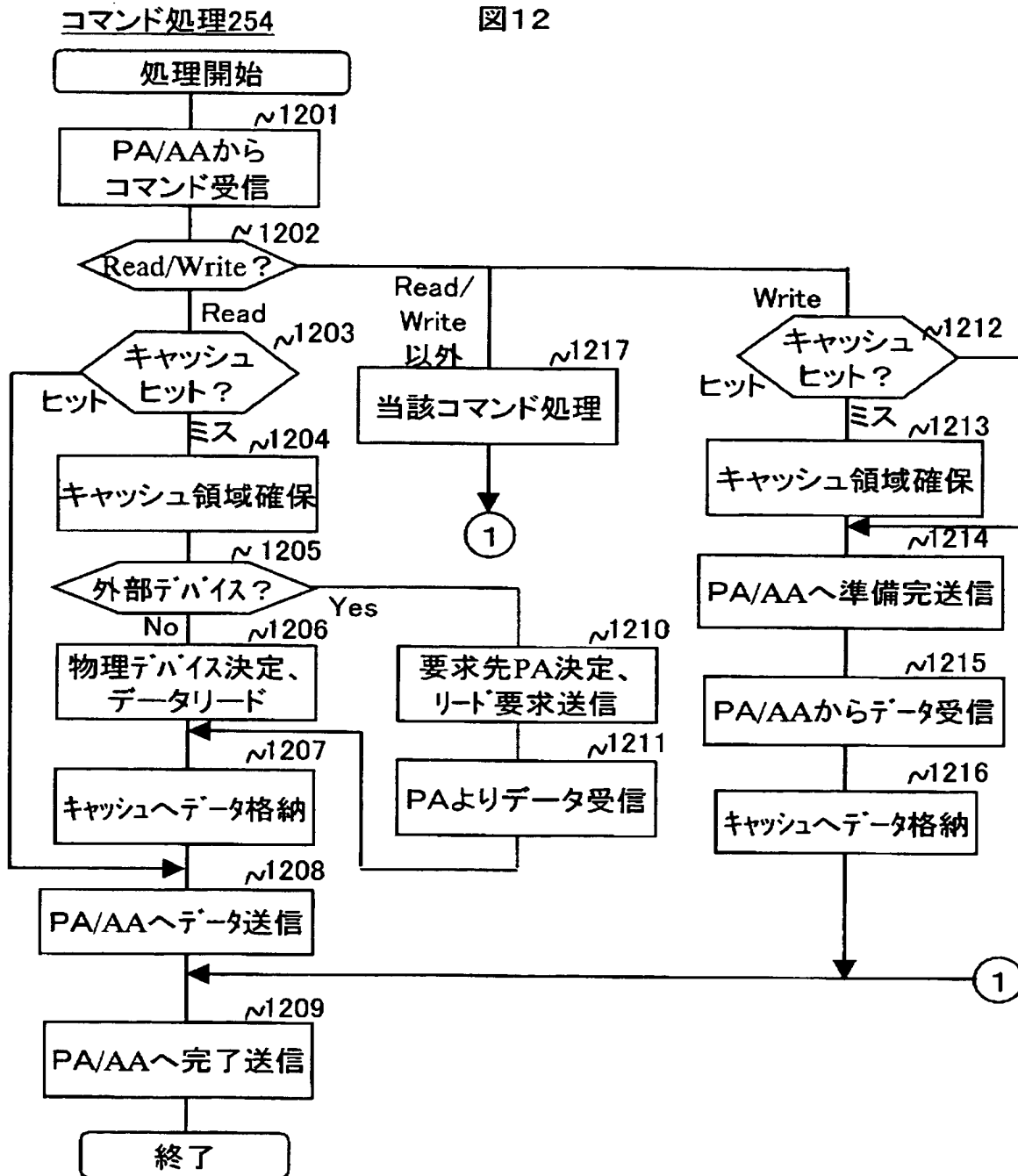
図 10

LUパス定義処理252

【図 11】



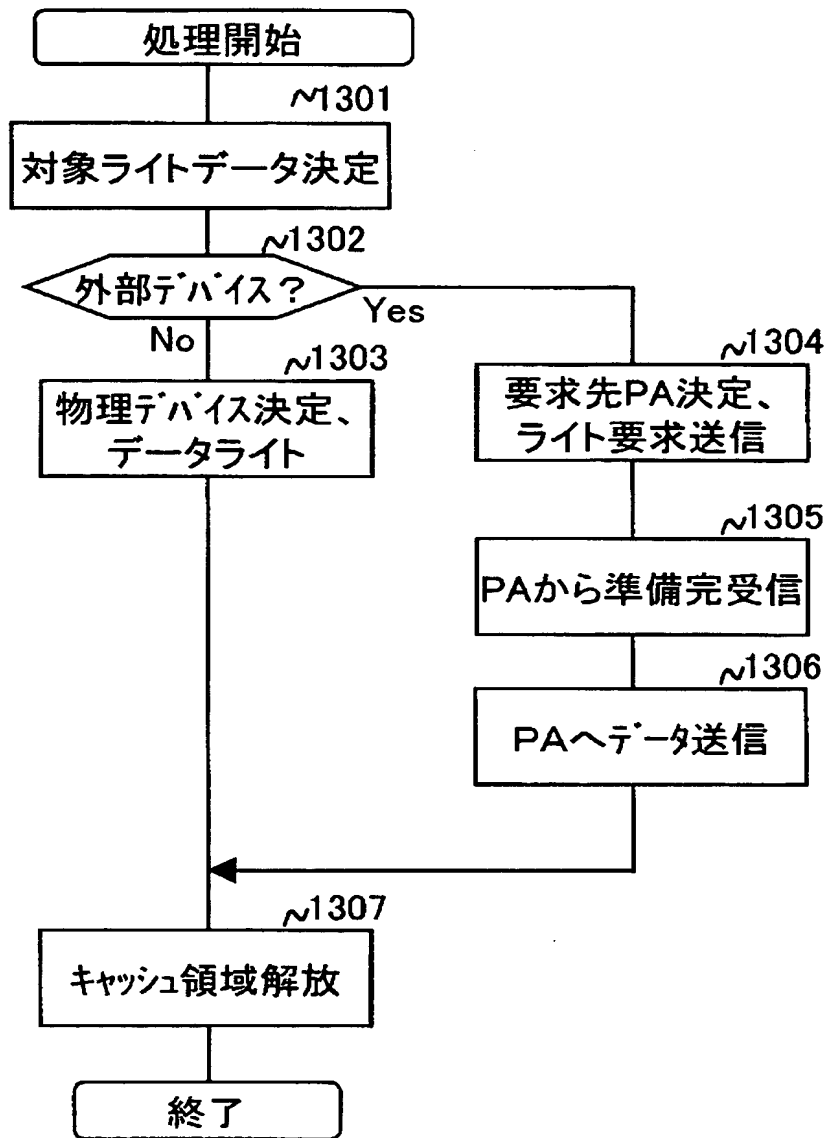
【図 12】



【図13】

ライトアフタ処理257

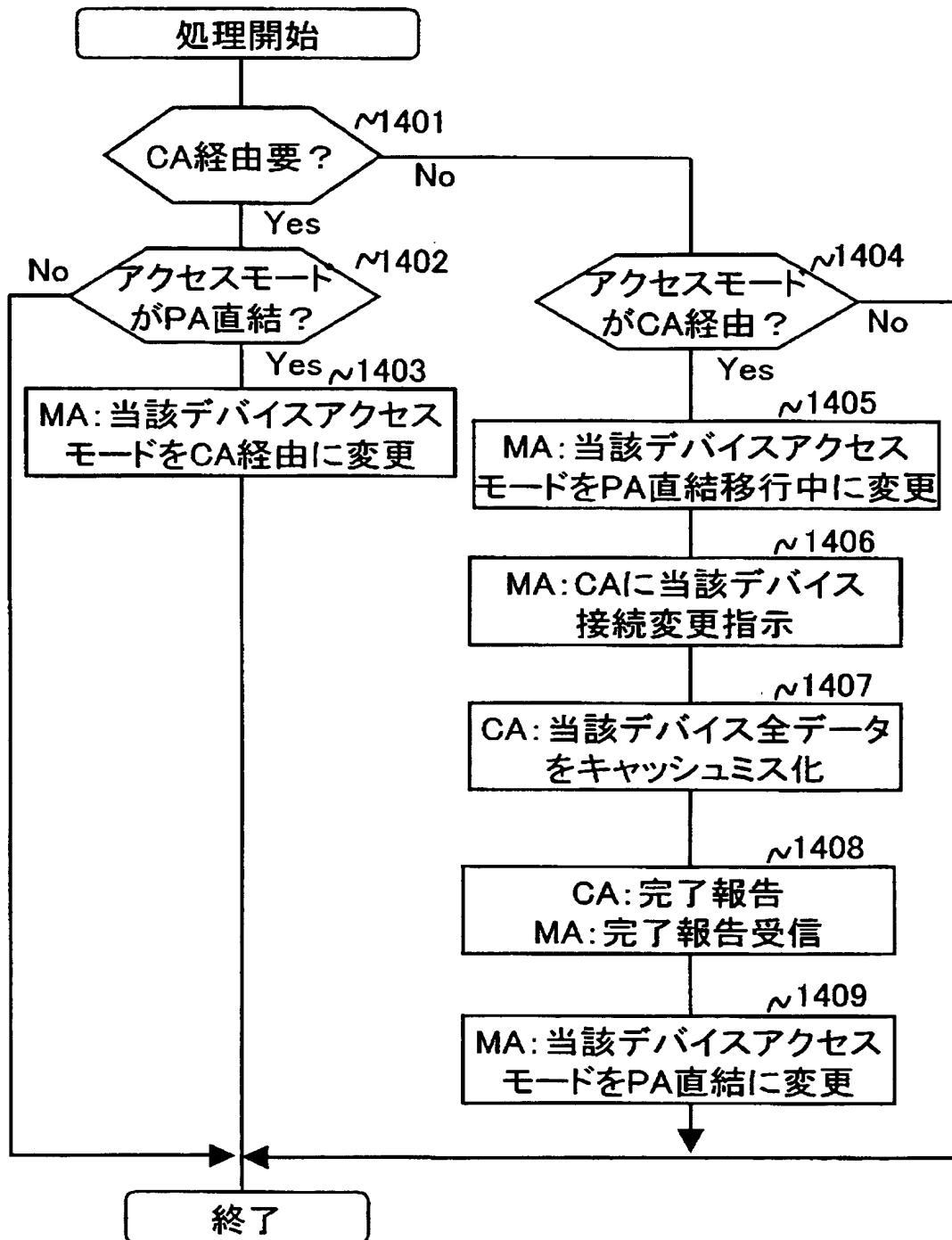
図13



【図 14】

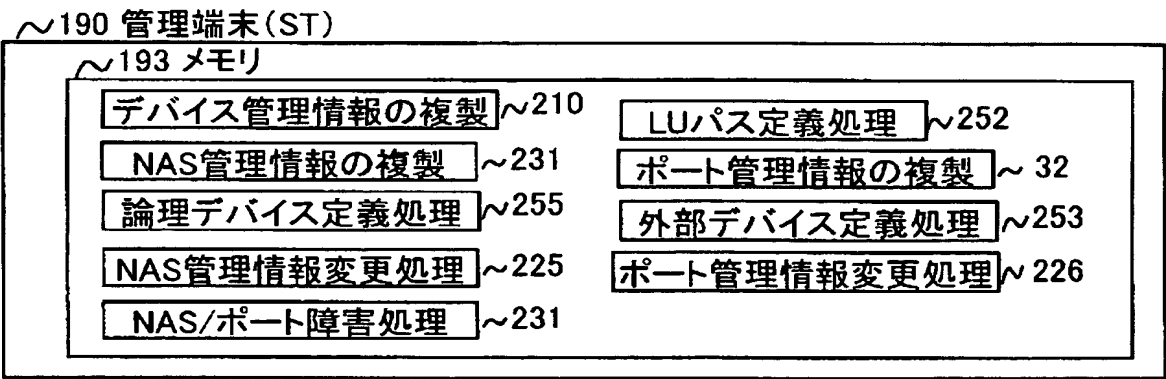
図 14

外部デバイス接続変更処理256



【図 1 5】

図15



【図 1 6】

NAS管理情報 223

図16

AA#0の 構成情報	AA番号	~1601
	IPアドレス、ディレクトリ名、対応CA番号、上位論理デバイス番号／外部デバイス番号のリスト	~1602
	状態	~1603
	アクセスモード	~1604
	代替AA番号	~1606
	⋮	

【図 1 7】

ポート管理情報 224

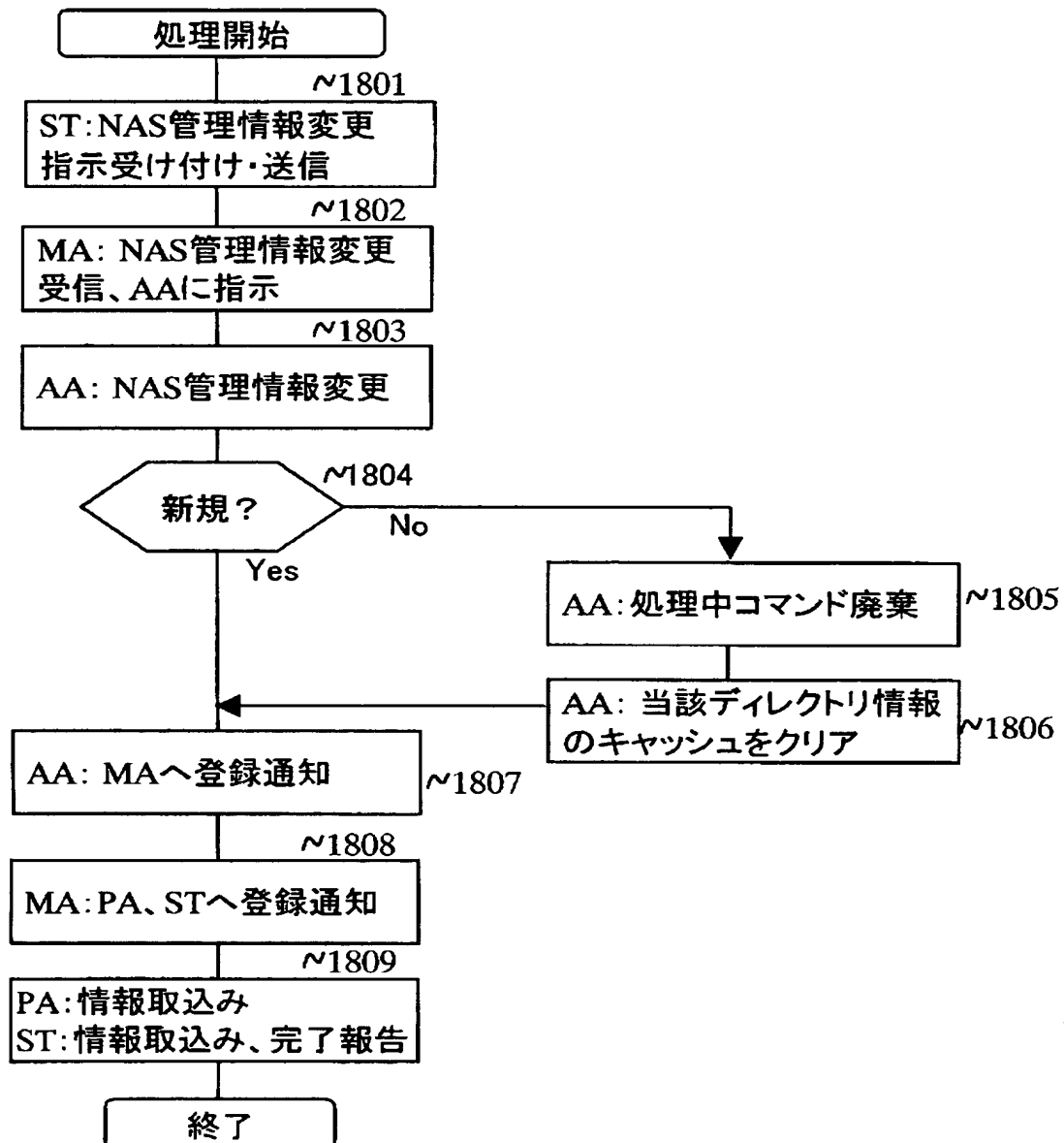
図17

ポート番号 #0
の構成情報

ポート番号	h1701
MACアドレス	h1702
IPアドレスのリスト	h1703
プロトコル	h1704
状態	h1705
アクセスモード	h1706
対応外部デバイス番号	h1707
代替ポート番号	h1708
⋮	

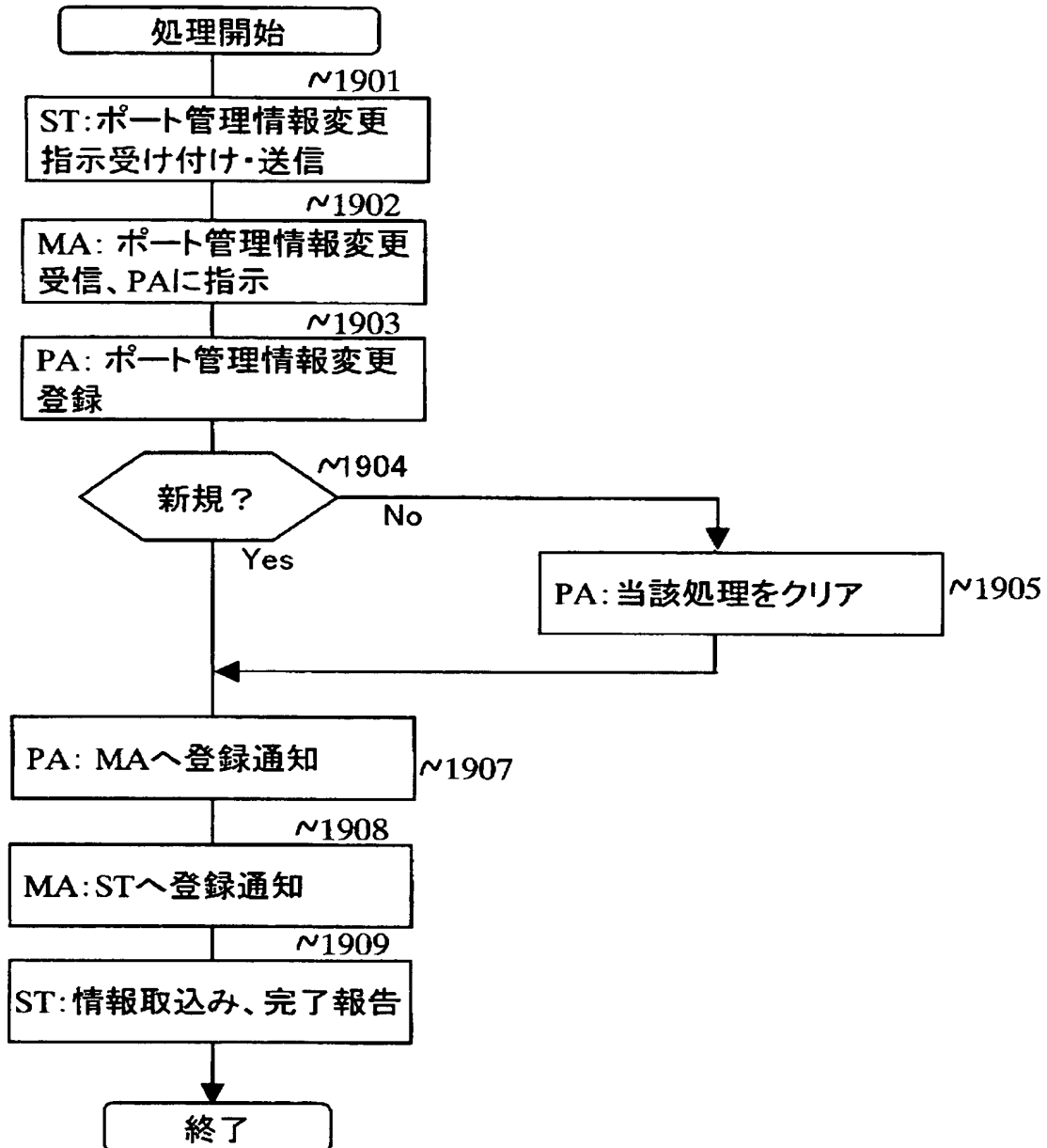
【図 18】

図18

NAS管理情報変更処理225

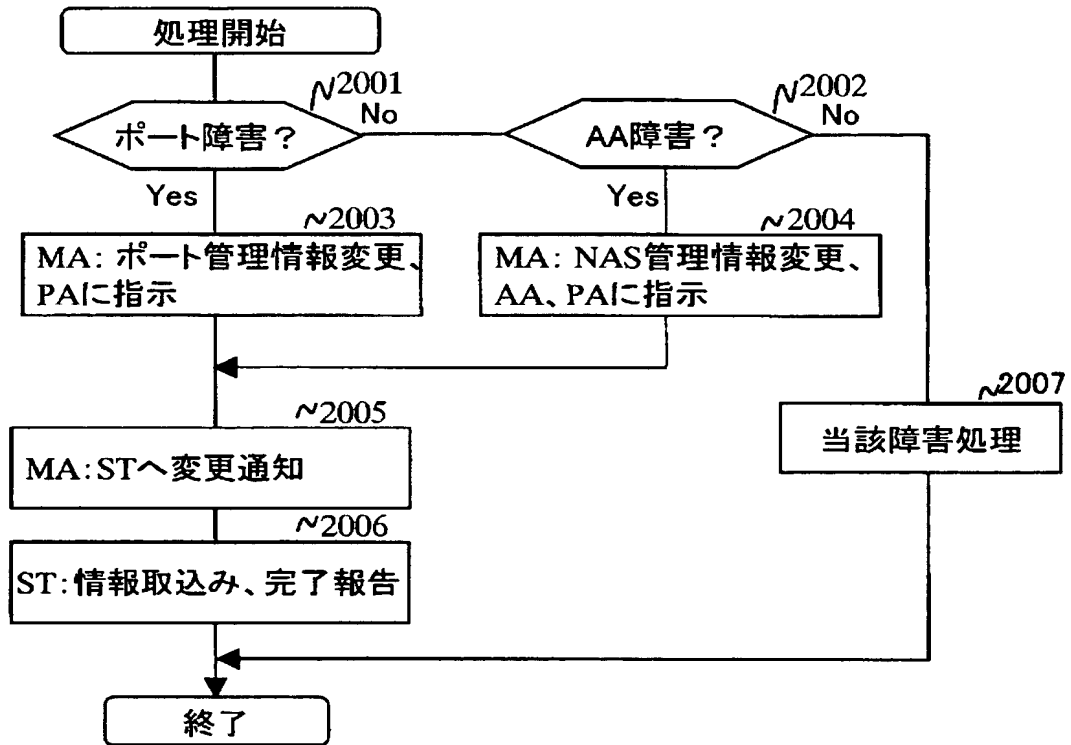
【図 19】

図19

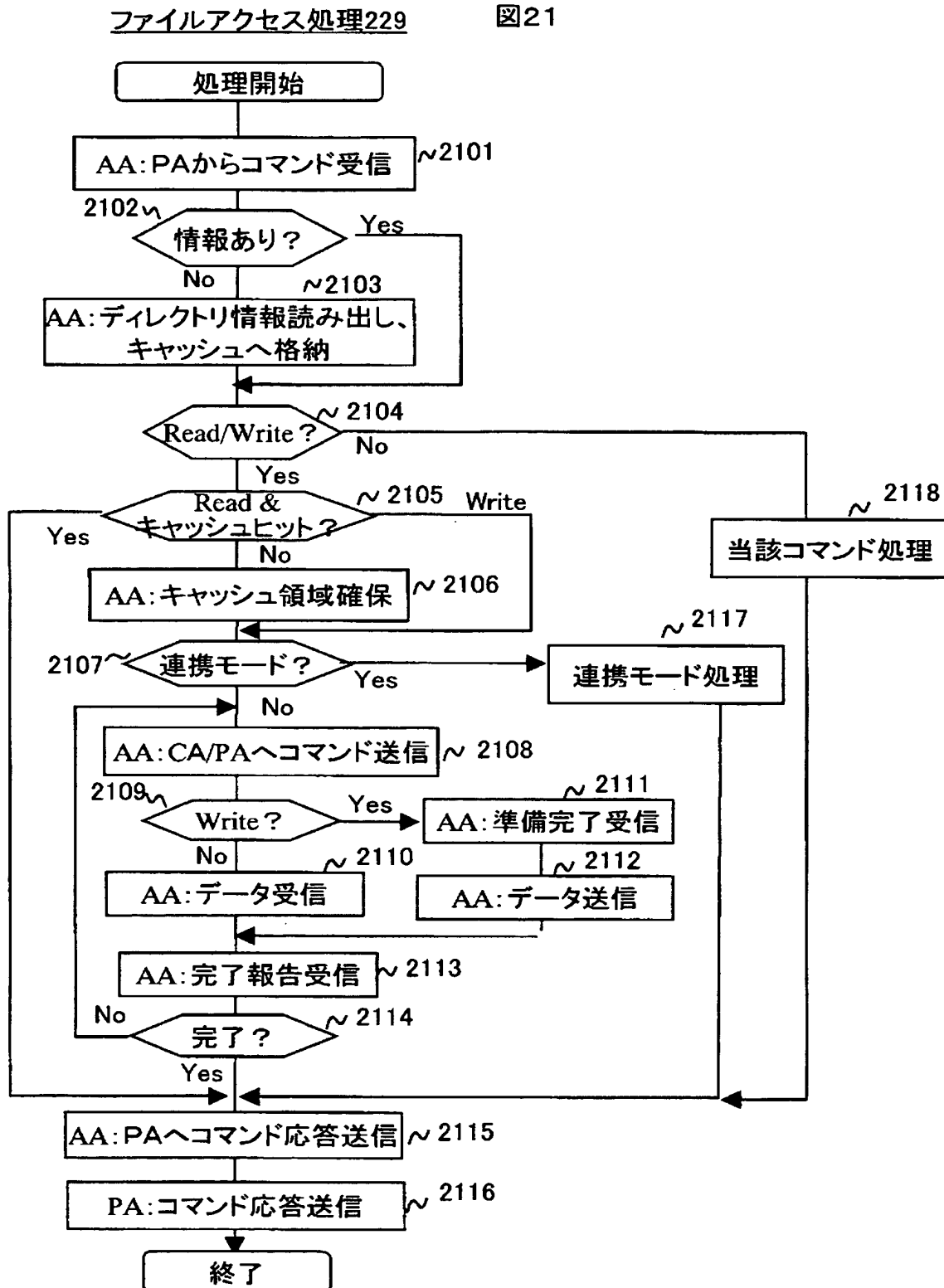
ポート管理情報変更処理226

【図 20】

図20

NAS/ポート障害処理231

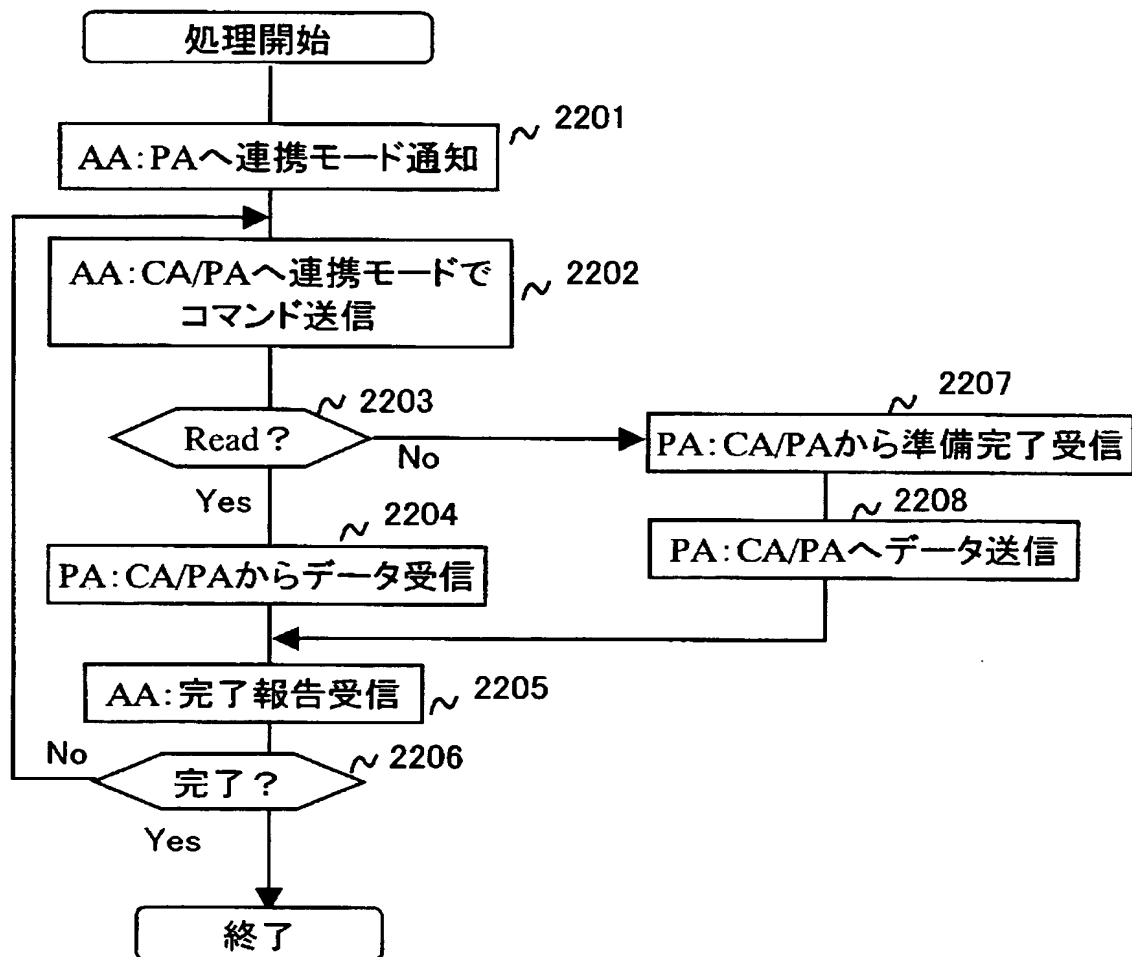
【図 21】



【図 22】

連携モード処理2118

図22



【書類名】 要約書**【要約】****【課題】**

NASサーバと当該NASサーバに付与されたネットワークアドレスが設定されるポートとの関係が1：1で一意的に決定されるためTCP/IP処理部またはNASサーバの処理性能がネックでレスポンスが低下する。

【解決手段】

ホストがコマンドを送付するアドレスが付与されるI/F部分とNASサーバ部とデバイスの処理を行うキャッシュ制御部と外部ストレージを接続するI/F部分を分離し、相互結合網170で接続することで、I/FとNASサーバの対応関係およびI/Fに付与されるアドレスとNASサーバとの対応関係を処理負荷等に応じ任意に変更する。

【選択図】 図1

認定・付加情報

特許出願の番号	特願 2 0 0 3 - 2 0 6 1 6 8
受付番号	5 0 3 0 1 3 0 0 0 4 2
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 5 年 8 月 7 日

< 認定情報・付加情報 >

【提出日】	平成 1 5 年 8 月 6 日
-------	------------------

特願 2 0 0 3 - 2 0 6 1 6 8

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 5 1 0 8]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所